

The weight of water: Benchmarking for public water services

Environment and Planning A
2016, Vol. 48(11) 2181–2200

© The Author(s) 2016

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0308518X16654913

epn.sagepub.com



David A McDonald

Global Development Studies, Queen's University, Kingston, Ontario, Canada

Abstract

The use of benchmarking and performance indicators to evaluate and compare water operators is a relatively new phenomenon. It has been taking place in the private sector since the 1970s but only migrated to public services over the past two decades. It is now widespread in the water sector, but there are emerging concerns about its commercial bias and relatively undemocratic processes. This paper reviews the history of benchmarking in the water sector, discusses arguments for and against its use, and proposes an alternative performance evaluation framework that may help to better account for universality, sustainability and democratic forms of governance, particularly with public water operators in low-income settings in the global South. I argue that shared forms of performance measurement can be useful but only if they are more explicit about recognizing local difference and if they become better at promoting public awareness and advancing equity. The paper also asks why existing water benchmarking systems do not explicitly differentiate between public and private water operators, and proposes indicators that may help promote non-commercialized forms of public water services. The proposals are necessarily tentative and preliminary – calling for more empirical and theoretical research on the topic – but do offer concrete alternative benchmarking possibilities for further debate and exploration.

Keywords

Benchmarking, water, alternatives, public, commercialization

Corresponding author:

David A McDonald, Global Development Studies, Queen's University, Mac-Corry Hall, Room A407, Kingston, Ontario, Canada K7L 3N6.

Email: dm23@queensu.ca

“We are all benchmarkers now.”

Jacques Santer, former President of the European Commission. (quoted in Sisson et al., 2003: 16)

Introduction

Measuring and comparing the performance of water operators via benchmarking is a relatively new practice. Having been popularized in the private sector in the 1970s it migrated to public services in the 1990s and is now widely practiced in the water sector.

The growth of benchmarking has sparked a vibrant debate about the choice of performance indicators and the challenges of managing complex and technocratic evaluation processes, but relatively little attention has been paid to its ideological and institutional foundations in water services. Nor has there been much analysis of how water benchmarking has been adopted in low-income countries in the South. The latter is of particular significance given recent efforts to (competitively) export performance evaluation frameworks from Europe and North America to countries in Asia, Africa and Latin America.

This paper offers a brief history of benchmarking in the water sector, followed by a presentation of three critiques of benchmarking practice that have been under-theorized and under-debated in the literature to date. I look specifically at the arguments that benchmarking tends to promote commercialization, its relatively undemocratic decision-making processes, and its predisposition to imposing universalistic (and Eurocentric) values on a heterogeneous water sector. I develop these arguments from a disparate set of literatures in an attempt to weave together a hitherto unconnected set of critiques of benchmarking in the water sector.

My intent is not to prove or disprove these various theories on benchmarking here. There is simply not enough empirical research on the topic to know conclusively how it shapes water service outcomes, particularly on questions of equity in countries in the South (despite its widespread application). There are, however, sufficiently indicative trends around its commercializing influences and bureaucratic procedures to raise valid questions about its impact on water service provision. As such, the paper is as much a call for further study as it is a cautionary tale about the possible influences of this fast-growing policy trend.

Nor is my intent to reject benchmarking. As alarming as the criticisms may be, benchmarking is not *inherently* commercial, undemocratic or homogenizing. It can, and should, be harnessed in ways that promote equitable forms of public water services and which democratize decision making procedures. Unlike critics that reject all forms of universal measurement, I conclude the paper with some preliminary proposals for what an alternative benchmarking process could look like in the hopes of stimulating further discussion and debate on both theory and practice.

The rise (and rise) of benchmarking

Modern benchmarking originated with private industry in the 1950s and was popularized by the Xerox Corporation in the 1970s in the United States in response to what the company saw as a rising competitive threat from Japanese technological firms (Camp, 1989; Levy and Ronco, 2012; Sisson et al., 2003). It is grounded in the use of quantifiable performance indicators for analyzing the internal activities of individual organizations, but takes on comparative yardstick characteristics when used to compare performance across organizations as well as across place and time (Pidd, 2012). For benchmarking to work, therefore, organizations must all (and always) use the same performance metrics.

By the mid-1990s “almost four out of five companies in Europe, North America and South East Asia were reported to be using benchmarking”, prompting then-president of the

European Commission, Jacques Santer, to claim that “we are all benchmarkers now” (Sisson et al., 2003: 16). Enthusiasm for benchmarking in the water sector developed soon after – notably along with the rise of New Public Management and the push for privatization. In fact, it was the privatization of water in the United Kingdom that was arguably the single most important catalyst in the adoption of benchmarking by water agencies, with the creation of the Water Services Regulation Authority for England and Wales (OFWAT) in 1989 intended to quantitatively track the effects of privatization over time.

Water regulators and policy makers in other countries took note, with the benchmarking process quickly spreading to other sectors, such as education, electricity and health (Alegre et al., 2006; Jackson and Lund, 2000; Jamasb and Pollitt, 2001; Wait and Nolte, 2005). In the United States, the 1993 Government Performance and Results Act “precipitated a virtual orgy of measurement” in the public sector (Pollitt, 2000: 120), with an associated proliferation of academic work on the topic. Dozens of books, hundreds of articles, and entire academic and managerial journals are now dedicated to performance evaluation and benchmarking in a wide range of sectors (although remarkably little of it engages critically with its theoretical presumptions or institutional foundations (Francis and Holloway, 2007)).

Other key milestones in water benchmarking include the establishment of the International Benchmarking Network for Water and Sanitation Utilities (IBNET) by the World Bank in 1996, the formation of task groups within the International Water Association (IWA) to explore ‘best practices’ in the late 1990s, the publication of the International Organization for Standardization’s (ISO) 24510 series on ‘Activities relating to drinking water and wastewater services’ in 2007, and the establishment of an annual IWA-sponsored international conference on performance measurements from 2008 onwards (Alegre et al., 2008; Bowerman et al., 2002; Cabrera et al., 2010, 2011; Danilenko et al., 2014; Parena et al., 2002).

There are now dozens of national water benchmarking associations and a growing number of regional groups. European water operators have been particularly active in this regard, with several established regional benchmarking programmes in place (e.g. the European Benchmarking Cooperation (EBC) (largely Dutch and Scandinavian) and Aquabench (largely German)). Additional programmes are being developed by the Organisation of Economic Co-operation and Development (OECD), the European umbrella association for national water federations (Eureau) and the European Commission (which, unlike other models mentioned here may be mandatory since the Commission has full legislative powers).

By contrast, there are very few national – and virtually no regional – benchmarking associations dedicated to water services in Asia, Africa or Latin America (Berg, 2013; Berg and Corton, 2007; Corton and Berg, 2009; Global Water Operators Partnership Alliance (GWOPA), 2009). Performance evaluation is widely practiced in these regions, but benchmarking systems are generally imported (some would say imposed) from elsewhere in the world, often at the request of international financial institutions, and sometimes as a loan conditionality. The World Bank, for example, makes its IBNET evaluation system available to national and local governments that want to “receive financing for capital improvements” (Van den Berg and Danilenko, 2011: 4; see also Water and Sanitation Programme (WSP), 2010).

There also appears to be an (unstated) race to expand these Northern-based benchmarking systems to countries in the South. IBNET has declared itself to be “the first global benchmarking standard... providing a global yardstick against which utilities and policy makers can measure their performance” (Van den Berg and Danilenko, 2011: 2). Not to be outdone, AquaRating (a proprietary, for-profit benchmarking agency developed

by the Inter-American Development Bank, with a focus on Latin America) claims to be “the only international system that facilitates an objective and comprehensive assessment of a utility’s performance, making it a pioneering tool in the global water sector.”¹ EBC meanwhile is piloting its benchmarking model in East Africa (on a not-for-profit basis).²

In short, performance measurement in the water sector is growing, and benchmarking frameworks are multiplying. Like it or not, “performance measurement is here to stay” (Poister, 2003: 21) and “we will be hearing many new benchmarking stories in the future” (Cabrera, 2008: 7).

Benchmarking consensus

Despite rivalry in the benchmarking community there is broad consensus amongst its advocates as to how and why it should be done. It is seen to enhance (and enforce) transparency and accountability amongst water operators by making performance data available to the general public and allowing people to compare their water operator with utilities in other jurisdictions. It is also seen to create opportunities for public participation in decision making by allowing “customer groups and NGOs to exercise ‘voice’ in an informed way”,³ as well as contributing to “consensus-based global solutions” for achieving the UN’s Millennium Development Goals (ISO, 2012: 2). Finally, benchmarking is seen to provide a common language and system for identifying and understanding ‘best practice’, offering water operators the quantitative targets and concrete tools for getting there.

There is also broad consensus as to what gets measured. Although every benchmarking system has its own unique features, most draw heavily (if not entirely) on the performance metrics established by the ISO 24500 series, with more than 260 standards for water quality, and another 550 related to water services more broadly (ISO, 2012: 3–4). The IWA, meanwhile, acts as a *de facto* gatekeeper of performance criteria by dint of its role as a global umbrella organization for water operators (public and private), especially since the creation of its “Specialist Group on Benchmarking and Performance Assessment” in 2010.⁴ As a case in point, the EBC claims to be “fully aligned” with IWA and ISO benchmarking protocol, using these indicators as “repositories” of performance criteria “for reasons of standardisation” (EBC, 2014: 5). In other words, there may be a growing number of benchmarking organizations, and they appear to be in some form of competition with one another, but they operate on essentially the same principles and use mostly the same performance criteria.

It is not possible to list the 260+ plus indicators that commonly make up evaluation systems, but the following list is indicative of the type and range of measurements that take place: number of water and sanitation workers per 1,000 connections; length of transmission and distribution mains renovated; percentages of unaccounted-for water; number of complaints due to water supply interruptions; average household incomes; book value of fixed assets; number of microbiological tests carried out for various chemicals; volume of electricity consumed; total capitalized cost of self-constructed assets; per capita consumption of water; number of mains failures; average time to complete repairs; and average water charges for non-residential consumption.

Measuring is difficult

Consensus on how and why to do benchmarking aside, the actual collection of data is anything but straightforward. Even the best-trained and well-resourced of water operators find it difficult to stay on top of what has become an onerous process of data collection, and

the situation is much worse for under-capacitated utilities in low-income countries where it can be difficult to gather even the most basic of statistics (Alegre et al., 2008; Berg and Corton, 2007).

Once collected, analyses and comparisons of data are no simple matter either. There are intense debates within the benchmarking community – albeit largely impenetrable to non-specialists – about the use of metric versus process benchmarking (Cabrera, 2008), partial indicators versus total or combined factor analysis, and differentiated statistical methods for analyzing complex information with linear programming methods such as data envelope analysis and stochastic frontier analysis (Abbott and Cohen, 2009; Parsons, 2002). Different clusters of indicators are also used, with EBC, for example, employing five key “performance areas” (EBC, 2014: 7), IBNET using 13 categories⁵ and AquaRating (2014a) using an entirely different set of eight “rating areas”, all of which make direct comparisons across different benchmarking frameworks extremely difficult.

As a result, abstract and highly technical forms of measurement can take on a life of their own, disconnected from larger utility objectives yet shaping policy and practice. Externally defined performance indicators can skew in-house priorities, stifling local identity and resulting in an abstract fixation on outside procedural norms: “At its worst, instead of being a force for change, benchmarking can put a stop to serious analysis of problems and/or experimentation with [locally-defined] innovative solutions” (Sisson et al., 2003: 23). It can also result in an anxious feeling of “keeping up with the Joneses” (Valcik et al., 2012), even when high-level benchmarking may not be feasible or appropriate for a local water operator. In some cases, benchmarking can cause a situation where water operators are measuring for measuring’s sake, with benchmarking becoming an end, rather than a means, to improved water services (Pidd, 2012).

The truism that “What gets measured gets done” (Osborne and Gaebler, 1992: 146) can aggravate the situation further, focusing attention on criteria that may not be the most relevant to a particular water operator or its end users. As Francis and Holloway (2007: 177) note, “[b]enchmarking is no more immune to the GIGO (garbage in, garbage out) principle than any other performance management system”.

Criticisms of benchmarking

For proponents of benchmarking these challenges are not deemed fatal to the measurement enterprise. Nor have these tensions altered the underlying principles of, or enthusiasm for, standardized performance evaluations. More radical criticisms do exist, but the literature is surprisingly thin and under-theorized given how pervasive the practice of benchmarking has become (Francis and Holloway, 2007: 172).

I outline three critiques below, drawing as much on the work of those who celebrate benchmarking as I do on its detractors, weaving together an admittedly eclectic set of theoretical concerns. The first argument is that current benchmarking practices promote commercialization. The second is that benchmarking processes are not very transparent or democratic. The final critique is that benchmarking imposes a singular worldview on a diverse set of international water beliefs and practices, marginalizing alternative perspectives on water services, particularly in the South.

Promoting commercialization

For arguments related to the commercializing tendencies of benchmarking one need look no further than the literature from benchmarking advocates themselves. For many proponents,

benchmarking is an intentional and explicit tool for commercialization, offering an “alternative to market forces” (van Helden and Tillema, 2005: 339), introducing competitive pressures to an inherently monopolized sector. With little or no possibility of direct competition, benchmarking is seen to simulate and stimulate market behaviour, pushing water operators to lower costs and improve services. According to IBNET:

Inter-utility performance comparison is needed in the water and sanitation sector because the sector offers limited scope for direct competition. Firms operating in competitive markets are under constant pressure to outperform each other. Water utilities are often sheltered from this pressure, and it frequently shows: some utilities are on a sustained improvement track, but many others keep falling further behind best practice. This matters, because a well-run water utility is essential to people’s lives. Only the most efficient, financially viable utilities are able to respond to urban growth, connect the poor, and improve wastewater disposal practices.⁶

Financial viability is key – a metric that many see as “the ultimate value of utility benchmarking” (Van den Berg and Danilenko, 2011: 8). Improving cost recovery and reducing per unit expenses have become the gold standard of measurement in the water sector, with financial indicators such as ‘percentage of unpaid-for water’ or ‘number of employees per 1,000 connections’ sometimes taking proxy for overall performance. Financial criteria are not the only measurement standards but they constitute a large proportion of benchmarking indicators and attract a disproportionate share of attention from policy-makers and funders, reflected in part by the massive literature on financial outcomes in the water sector and support from development institutions such as the World Bank for improving cost recovery (Alexander, 2005; Breen and Doyle, 2013).

Some benchmarking advocates see this emphasis on financial performance as a way to promote liberalization and privatization, forcing public water utilities to make their fiscal data available for corporate review and helping to identify markets for potential private investment. According to IBNET, “[p]rivate investors interested in expanding their interests in the water and wastewater sector can use [benchmarking] to carry out an initial screening of potential target utilities”, helping them to “pinpoint those with revenue-generating potential”⁷ and to “identify viable markets and opportunities for creating value” (Van den Berg and Danilenko, 2011: 4).

As such, involvement in the development and implementation of benchmarking criteria is often touted as an opportunity to be at the cutting edge of competitive water markets. The ISO encourages firms to “get involved” in the development of benchmarking because it “can bring significant advantages to your business” via the establishment of standards that fit with an organization’s strengths or by giving advance notice of what the market will expect. By creating market-friendly indicators, benchmarking can help firms “access new markets” and “facilitate free and fair global trade.”⁸

Not surprisingly, then, many international benchmarking organizations are composed heavily of multinational corporations keen to shape global standards across a wide swath of sectors, from environmental sustainability to corporate governance (Clapp, 1998; Nadvi and Waltring, 2004; Prakash and Potoski, 2006; Stevenson and Barnes, 2001). The ISO has come under particular fire in this regard, with critics arguing that most of its committee work is conducted in a handful of countries in the North and dominated by large private firms, making it little more than a “corporate private regime” (Haufler, 2004: 126). Proposals by the European Commission to develop an EU-wide benchmarking system for water have received similar criticism, with Aqua Publica Europea (APE) noting “the unbalanced nature of [the planning group’s] membership. . . [T]he public sector is scarcely represented, if at all” (APE, 2014: 6).

The result is performance evaluation systems in the water sector that “strongly motivate [operators] to be efficient and innovative, mitigating their operating costs and expenses” (Marques and Simões, 2010: 15; see also Bowerman and Ball, 2000). In this respect, it can be argued that benchmarking has contributed to the naturalization of financial efficiency in the water sector, converting economic goals into “neutral facts” that validate and reproduce otherwise contestable and diverse aims (Boelens and Vos, 2012: 18). Benchmarking can, as with any other “absolutized efficiency calculus”, be used “as a weapon to suppress contending social groups, their social analyses, and their programs for social change” (Wolff, 2002: 3).

None of this is to say that benchmarking is inherently commercial. Nor does it automatically bind water operators to their financial bottom line. Many water managers and policy makers are intensely aware of – and often opposed to – commercializing influences in the water sector. Some may actively resist marketization. There is also little evidence that benchmarking actually results in improved financial discipline (Braadbaart, 2007). Indeed, many water managers dislike change, preferring to do things ‘the old way’, making *any* kind of new benchmarking systems difficult to introduce (including my proposals, below, for an alternative benchmarking system).

But given the influence of organizations such as the World Bank in promoting financialization and marketization in the water sector (Bayliss, 2014), and their stated intent of using benchmarking as a market substitute, there are ample grounds for concern around the impacts of this growing trend on water operator behaviours and outcomes. The fact that none of the major water benchmarking systems mentioned above make any distinction between public and private water operators only serves to heighten these apprehensions.

Benchmarking as undemocratic

A second concern with benchmarking is that it tends to be done in a top-down, undemocratic manner, excluding water users and lower-level employees. Because of the highly technical nature of performance evaluation most benchmarking systems make little (if any) effort to involve non-experts. Most water users are probably unaware of benchmarking processes and the 260+ performance indicators that go into them (let alone the advanced statistical analyses that accompany their assessment), essentially rendering the benchmarking process opaque. Nor is the average water user likely to be tempted to participate in such a highly specialized and mechanical procedure.

Large benchmarking organizations have done little to make these processes more transparent or inclusive of the broader public. Although an increasing number of benchmarking reports are made available for public review, “simply publishing the results does not necessarily equate to more transparency” (APE, 2014: 2). Benchmarking reports are technocratic in method and difficult to decipher, and are seldom even available to the general public. As a case in point, access to IWA and ISO benchmarking models requires an expensive membership, while AquaRating is a proprietary framework. Little wonder that making sense of benchmarking data is the domain of a relative few.

Some critics see in these actions a form of neoliberal governmentality, with monopoly control over a “governing technology that seeks to facilitate the self-governing capacities of individuals and/or organizations through the production of a normalizing knowledge” (Triantafyllou, 2007: 836). Insofar as benchmarking is an attempt to create an “international consensus on solutions to water issues” (ISO, 2012: 4), it is seen by some as a mechanism for “producing truth...steering social behavior and giving normative

meaning to particular water practices of particular water user groups... convincing not only the actors who have to apply these concepts but also the creators themselves” (Boelens and Vos, 2012: 18; see also Vos and Boelens, 2014).

Benchmarking systems, it is argued, must be seen within the “social and power relationships” in which they are embedded, with planners and managers “often not aware of their value-loadedness, convinced that they provide objective advice that should be adopted by policy-makers and politicians” (Boelens and Vos, 2012: 24). In this regard, benchmarking can validate concepts of success and reinforce ‘best practices’ that advance particular agendas. The truism cited earlier – “What gets measured gets done” – can result in an unremitting cycle of investment and policy emphasis that advances the commodification of water and justifies inequities, particularly in countries in the South where opportunities for democratic input and alternative voices can be very limited.

Once again, this is not a necessary feature of benchmarking, or an intentional outcome on the part of water managers or benchmarking designers – some of whom are expressing concern with its exclusive nature (APE, 2014). But the reality of performance evaluations is that they are largely impenetrable to the general public, while the lack of robust debate on this topic is itself indicative of the enigmatic nature of benchmarking in practice.

Benchmarking imposes universality

A third criticism of benchmarking is that universal performance criteria essentialize and homogenize water and the people that use it, running roughshod over cultural and political difference by imposing generalized standards on the world. The argument here is partly an epistemological one, questioning the very validity of objective and universal standards inherent in the notion of comparative performance evaluation. Critics argue that there are no constant, universal truths: “Democratic conceptions of the common good will always be partial and provisional, never universal or static... the common good can never be specified *a priori* as an input for the political system or as a static measure for the quality of governance” (Dahl and Soss, 2012: 31). For Zwarteveen and Boelens (2014: 151–152), “knowledge about water will always and necessarily be uncertain and provisional,” reminding us that we must “remain vigilant about the temptation to unequivocally use ‘science’ and the objectification it entails in dealing with water’s complexity.” For these critics, it is simply not possible to measure and compare performance in the water sector across place and time because no two communities are ever alike, while public conceptions of water are constantly changing.

The implication here is that benchmarking can smother the very improvements it seeks to promote. In attempting to impose standardized measures of value, measurement processes serve to “fix the public in place, rendering it static in a manner that contradicts the promise of an evolving constitutive democratic process” (Dahl and Soss, 2012: 22). What is required instead is a focus on the creation of space and resources for “deliberative processes that... allow new conceptions of public identities and interests to emerge” (Dahl and Soss, 2012: 22–23). Rather than relying on notions of a public that is “always a pre-existing collectivity... that can be identified, addressed and moved to action,” we should see publics and the values they hold as “entities that are always mediated and always emergent” (Mahoney and Clarke, 2013: 932). Relaxing the demand for universal standards “usefully opens the door to accepting diverse and plural knowledges about processes of water-related change – including those based on the experiences and knowledge of people who live in changing environments” (Zwarteveen and Boelens, 2014: 151–152). Benchmarking is seen to disqualify and marginalize alternative forms of water management and valuation (particularly ‘pre-modern’ forms), taking on the

hue of colonial practice “imposed from an outside position as a way to bound or police governance” (Dahl and Soss, 2012: 31).

There are also concerns that benchmarking standards have been largely developed by institutions and water operators in Europe and North America, with insufficient attention being paid to the realities of water providers in the global South. Cultural, political, economic and environmental differences may make some performance criteria irrelevant or inappropriate, while the sheer cost of benchmarking can make it impossible or irresponsible to do in full. Even within the European Union there are concerns about the suitability and feasibility of universal benchmarking models, with Aqua Publica Europea arguing that “many operators in rural or less favoured areas face cultural and economic difficulties in participating in benchmarking exercises” (APE, 2014: 6). This concern is magnified many times over in Africa, Asia and Latin America.

In search of alternatives

Given these critiques, should benchmarking in the water sector be abandoned? Yes and no. To the extent that benchmarking intensifies market pressures, excludes the general public and imposes inappropriate goals on diverse water practices it can be argued that it is not suitable for many places. If nothing else, it is crucial that there be more research and a more robust debate on how benchmarking plays itself out in practice, particularly in the South.

But I am equally convinced that benchmarking of some kind is essential to the advancement of ‘good’ (as opposed to ‘best’) water services. It may be problematic to develop universal standards, but we need not throw out the measurement baby with the proverbial bath water. It is possible to develop alternative forms of measurement and comparison that mitigate (if not resolve) the tensions outlined above and which allow for more democratic, less commodified and more explicitly *public* forms of water provision that can be used to compare and learn across place and time.

This will not be an easy political or technical task given the authority and resources of mainstream benchmarking organizations and their vested interests in existing models. So too will many water operators resist change, particularly those that have sunk resources and political capital into current benchmarking frameworks. Even managers and policy makers who share the concerns raised in this paper may find it difficult to change directions given the inertia of existing benchmarking systems and the time and energy required to shift analytical and operational gears. It has taken two decades of intense funding, lobbying and institutional support from major international organizations to get water benchmarking to where it is today. It is not going to change course overnight.

Nor will it be easy to sell the idea of revised benchmarking to those who are inherently skeptical of standardized, centralized evaluation (which includes many water activists pushing for more autonomous forms of water service provision). But a rejection of performance comparison vacates the possibility of more progressive and flexible notions of universality. Without some commonly agreed upon performance criteria how are we to know if a water operator is ‘successful’? How do we articulate demands for ‘improved performance’? How can we share ‘good’ practice across different places? Rejecting benchmarking altogether risks leaving this powerful tool in the hands of those who may, intentionally or not, seek to advance the commercialization and homogenization of water services.

These normative questions do not relieve us of the deep-seated philosophical and practical tensions between the universal and the particular – a theoretical debate well beyond the scope of this paper – but they do force us to ask whether it is necessary to develop standardized criteria for water services. On this point I take my cue from Harvey (2000:

94) who notes, with reference to analogous debates around the development of standardized codes for international human rights, “To turn our backs on such universals at this stage in our history... is to turn our backs on all manner of prospects for political action.” In other words, rejecting universal standards for water services runs the risk of not developing effective, concerted actions to deal with the hundreds of millions of people around the world with inadequate access to these basic amenities, possibly entrenching inequalities.

These questions also raise practical matters of *whose* universalisms or particularisms matter. On this point I take my cue from Timmermans and Berg (1997: 275) who – writing about universalized medical protocols – argue that standardization efforts need not require a central actor and that achieving universality should be seen as a “distributed activity” leading to “local universality”: “an ambiguous and pre-carious status of ‘universality’... [that] emerges from localized processes of negotiations and pre-existing institutional, infrastructural, and material relations”. In this case, universality has become a “non-transcendental term – no longer implying a rupture with the ‘local’, but transforming and emerging in and through it”.

My proposal is thus driven by a normative understanding that urgent change in the water sector requires a coordinated effort to develop an effective counter-narrative on performance evaluation, but that this vision can be done in a spirit of diffused effort. Such alternatives could offer standardized measurement principles and criteria while at the same time being supple enough to encourage local interpretation to pursue “diverse goals, such as equity, stabilization, and social and environmental sustainability” (Lefebvre and Vietorisz, 2007: 139–140). The objective is to construct a dialectical bridge between these universalisms and particularisms, while recognizing that generalizations are inherently fraught with cultural and political tensions that disallow easy comparisons and may be irreconcilable at times.

This is not to suggest that mainstream benchmarking has ignored the tension between universal norms and local realities. Indeed, much of the existing literature is at pains to highlight the need for ‘flexibility’ and ‘local difference’ (Baietti and Ginneken, 2006; Cabrera, 2008; Corton and Berg, 2009; Crotty, 2004). The problem with these efforts, as I have argued above, is that they tend to be couched in an efficiency calculus that overwhelms all other factors, marginalizing questions of equity and being largely uncritical of their institutional genesis.

Alternative evaluation methods can never resolve these tensions entirely. It is possible, however, to be more explicit about social, political and economic variability, especially in low-income countries in the South where equity is a major challenge. Recognizing radically different cultural interpretations of water, prioritizing equity in contexts of extreme poverty, and acknowledging diverse forms of water governance will require much more flexible systems of evaluation than those offered by current benchmarking models.

So too must alternative forms benchmarking be more inclusive, “promoting participatory governance mechanisms which, when coupled with transparency, empower citizens when it comes to decisions on the management of water resources” (APE, 2014: 2). Benchmarking processes and reporting will need to take better account of literacy, numeracy and analytical skills as well as barriers to participation along socio-economic lines.

So where do we begin? My suggestion is that we start with what we already have. There are many useful – I would argue essential – indicators already in place in current benchmarking models that should be retained. Metrics for water quality, evaluations of infrastructure repairs, frequency of emergency breakdowns and a host of other relatively objective criteria can be easily and usefully reproduced in new evaluation models.

What needs to change is (1) the ways in which we analyze some of these existing criteria, and (2) the addition of new indicators that better promote equity and ‘public’ water provision. An example of the former is unaccounted-for water. Instead of simply tracking

the number of leaky pipes and levels of non-payment, performance indicators could be used to ask whether water services are affordable for low-income households and whether adequate investments are being made in bulk in infrastructure in low-income neighbourhoods. A more layered understanding of these metrics could reduce harmful water cutoffs, improve the progressivity of tariffs and promote a more spatially equitable pattern of infrastructure investment.

On the second point, brand new indicators could be added, such as evaluations of worker health and safety, participatory decision-making mechanisms, and opportunities for female employees to move up the skills ladder. Such indicators could shed light on gender barriers to career progress, provide insights into opportunities for improved community engagement, and improve the quality of work for labourers – all important criteria that are largely ignored in current benchmarking systems.

It may also be wise to consider a much reduced number of indicators than current benchmarking systems employ – already a topic of debate in the benchmarking literature. Van der Steen (2011: 33), for example, suggests that the number of performance indicators should be in the 15–30 range. The City Blueprints for Water initiative uses 24 indicators: “a method that is practical, relatively simple, transparent, easy to communicate and understandable for decision-makers and the public in general,” taking about a week to conduct (van Leeuwen et al., 2012: 2180).

Pidd (2009) promotes the principle of “model simple, think complicated” (drawing on Little, 1970), arguing that benchmarking systems should be easy to understand, robust enough to allow counterintuitive results to emerge, simple to manage, adaptive to different situations, and easy to communicate. The challenge is to find a balance between the complex reality of water systems and the need for simplification in ways that “helps focus people’s minds” (Pidd, 2012: 75–76).

There are, of course, good reasons why performance evaluation has grown in complexity over the years. It is often because of the need for more progressive metrics of the kind that I am proposing here. However, *more* measurement does not necessarily mean *better* measurement. Just how many indicators are optimal is not clear, but the enormous volume of measurement indicators promoted by the ISO, and the byzantine statistical analyses that accompany them, are clearly too complex and too opaque for meaningful public debate, overwhelming even the best-resourced of water operators. An alternative benchmarking framework may therefore benefit from a more streamlined and popularly accessible approach.

I would, however, warn against false simplicity. AquaRating (2013: 5), for example, has developed a benchmarking model that provides a single evaluation score (a number between 1 and 100), but the calculations required to arrive at this figure are as complex as ever: “eight assessment areas, 27 assessment sub-areas, 113 assessment elements, 61 indicators, 99 variables, 52 practices groups, [and] 348 individual practices.” In other words, AquaRating’s ‘simple’ benchmarking system still requires experts and enormous time and resources while its results remain impenetrable to the average front-line worker and water user – presumably one of the reasons that they charge between US\$25,000 and US\$65,000 for an audit, plus “logistic expenses” (AquaRating, 2014b: 6).

A spider web evaluation

The alternative evaluation framework that I am proposing here (in very broad and preliminary terms) attempts to find a middle ground between the need for simplicity and the reality of complexity on the ground, while at the same time employing new and old

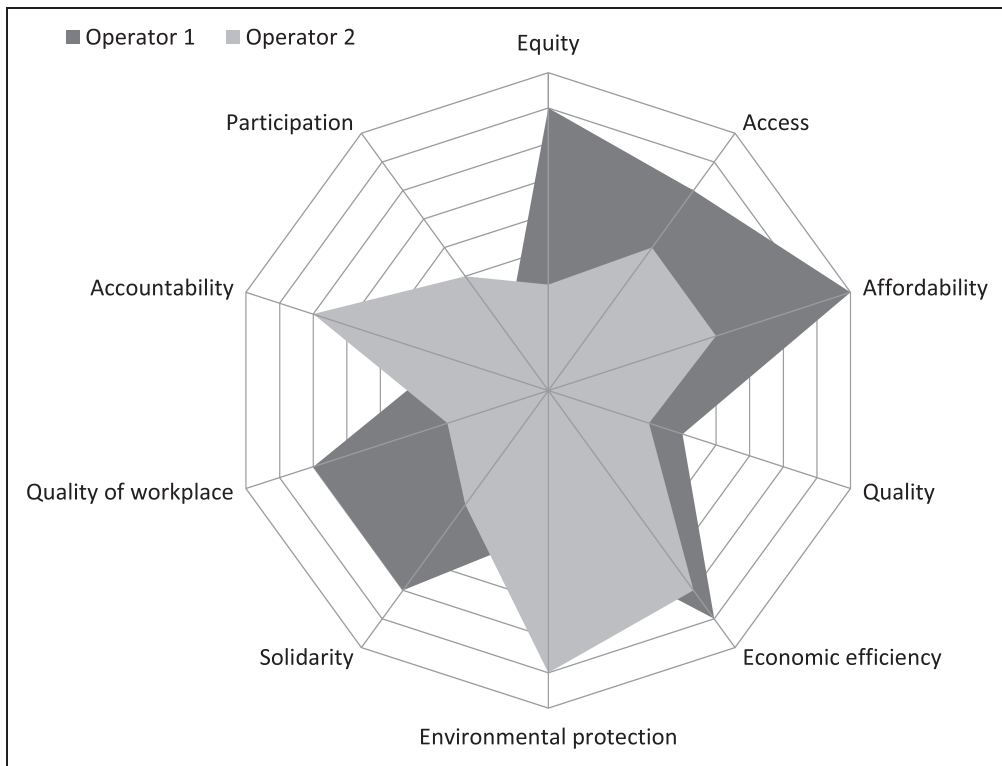


Figure 1. Spider diagram comparing public water operators.

evaluation methods. The intent is not to suggest a final, polished product, but rather to concretely demonstrate the possibility of a more user-friendly form of data collection and analysis that also visually represents results in ways that promote public engagement, accessibility and elasticity. In doing so I draw on the existing practice of clustering benchmarking categories (such as EBC's performance areas of "water quality, reliability, sustainability, service, and finance/efficiency" (EBC, 2014: 7)) and efforts by groups such as City Blueprints for Water to simplify and pictorialize their benchmarking system (van Leeuwen et al., 2012: 2180). The latter employs 'spider diagrams', which I have adopted here, although other visual representations could be equally effective.

Figure 1 is an example of such a benchmarking representation, comparing two hypothetical water operators. It employs nine overarching performance categories, each with a series of popularly understandable compound indicators. Visually, the longer the shaded area on each thread the better the service for that criteria, and the larger the overall shading on the web the better the overall performance of the operator.

Table 1, meanwhile, outlines the proposed indicators and metrics that could make up these valuations. Once again, these are not necessarily the best or the only benchmarking criteria – or the correct number of categories – but they serve to demonstrate what an alternative system might look like. Overall, the aim of these indicators is to prioritize what I have argued previously to be cornerstones of a 'public' service – universality, sustainability and democratic governance (McDonald and Ruiters, 2012). Financial resources are important but they should not overshadow other priorities, and should not be confined to a single sector. Water services are intricately linked to health and well-being

Table 1. Normative criteria for evaluating public service providers.

Normative category	Criteria	Definition	Examples of possible sub-criteria questions	Examples of possible measurement indicators
Universality	Access	Physical availability of the service at a convenient distance from user's dwelling	• Rural/urban divide? • Sufficient quantity? • Culturally acceptable service?	• Proportion of population with adequate access • Time-distance to service location • Hours/day that service is available
	Affordability	Prices that ensure economic accessibility for all	• Are poorer households disproportionately burdened? • Are programmes in place for cross-subsidy pricing? • Is affordability a legal obligation?	• Cost as percentage of household income • Disconnection rates • Levels of subsidization by region
Quality	Quality	Reliable, satisfactory services that create positive relations with end users	• Safe for all users? • Responsive to user needs? • Ongoing improvement mechanisms in place?	• Primary health outcomes • Level of service interruptions • Complaints by region
		Equality of opportunity to access quality services for all	• Equitable quantity of service across user groups? • Equitable quality of service across user groups? • Is equity formalized, legalized or institutionalized?	• Budget allocations by region • Levels of access by socially disadvantaged groups • Per capita consumption by region
Sustainability	Efficiency	Cost-effective use of resources to meet service mandates	• Are current infrastructure investments helping to meet the social goals of the service? • Is the capital intensity of investments appropriate? • Do short-term cost reductions undermine long-term efficiency gains?	• Financing as a proportion of overall operating costs • Cost per unit of service delivered by region • Employee turnover rates

(continued)

Table 1. Continued.

Normative category	Criteria	Definition	Examples of possible sub-criteria questions	Examples of possible measurement indicators
Governance	Environmental protection	Meeting current service mandates without compromising future resource needs or undermining cultural environmental norms	• Are programmes in place to reduce demand on natural resources? • Does the service provider respect different cultural understandings of resources? • Are climate change mitigation plans in place? • Does the model help to build a stronger 'public ethos' around services? • Is the service contributing to improvements in other sectors and at other levels of service delivery? • Does the service model explicitly oppose privatization and commercialization, with sufficient political support?	• Levels of renewable energy use • Quality of wastewater treatment • Rates of respiratory infection
	Solidarity	Cohesion among various producer and user groups and across sectors that builds economic, social and political commitment to a public service model	• Does the model help to build a stronger 'public ethos' around services? • Is the service contributing to improvements in other sectors and at other levels of service delivery? • Does the service model explicitly oppose privatization and commercialization, with sufficient political support?	• Formal cooperation agreements between different levels of government and sectors • Measurements of inter-sectoral impacts (e.g. sanitation extension reducing diarrhoeal burden?) • Legal mechanisms to prevent privatization
	Accountability	Obligation to account for activities, accept responsibility for them, and to disclose the results in a transparent manner; readily available to the public, and understandable.	• Are there clear operational mandates and policy positions? • Are there transparent capital and operating budgets? • Are mechanisms of accountability available at appropriate scales (local, national, regional)?	• Transparency of hiring processes • Access to mechanisms of accountability by region • % of documentation openly available and verifiable, in suitable languages and formats for all users
	Participation	Citizen involvement in policy making and implementation of service delivery	• Is participation at appropriate scales and sufficiently representative?	• Number of people participating in formalized mechanisms of participation

(continued)

Table 1. Continued.

Normative category	Criteria	Definition	Examples of possible sub-criteria questions	Examples of possible measurement indicators
Quality of workplace	A place of work that provides a safe environment, trust between employees and management, fairness, and a sensible workload that contributes to quality service delivery		• Are there adequate resources for participation by a diverse range of society (transportation, time off work, etc.)? • Is participation conducted in culturally appropriate ways? • Are there adequate numbers of workers to ensure service quality? • Are there mechanisms for workers/unions to participate in the operation, management or policy-making of the service? • Are there good feedback loops between front-line workers, managers and end-users of the service?	• Number of different processes of participation open to participation (policy making, budget decisions, etc.) • Availability of participation by region • Pay equity (job type, gender, race, ethnicity, etc.) • Availability of health and safety equipment • Access to training opportunities

across a range of services, requiring a multi-sectoral, multi-scalar and even multi-jurisdictional perspective.

The criteria and indicators proposed in Table 1 are suggestive of the kinds of evaluation questions and techniques that are aimed at equity and which go beyond the confines of water to look more broadly at the public goods that are produced (or not) from the provision of water services. The intent is to make more 'public' the otherwise narrow institutional and commercialized evaluation criteria found in existing benchmarking models. I have also introduced a category for 'solidarity', defined here as 'cohesion among various producer and user groups and across sectors that builds economic, social and political commitment to a public service model'. Evaluation of this category could include such questions as: Does the model help to build a stronger 'public ethos' around services? Is the service contributing to improvements in other sectors and at other levels of service delivery? These are not the only or most exhaustive questions to ask about the 'publicness' of a water operator, but they do serve as a counterpoint to the commercial bias of existing models and make explicit the need to evaluate a water provider's ability to address broad public good outcomes (as such, it would be useful to see how *private* water operators perform against a scorecard geared for *public* entities – an inverse of current benchmarking realities).

These proposed criteria are not *inherently* more attentive to flexibility or local difference. It is still possible to impose these suggested performance indicators in crudely universalistic ways, with little interest in local variation. The new criteria do, however, force water operators to grapple with provocative questions of inequality, accountability and publicness that can vary much more dramatically from place to place than mainstream benchmarking indicators. They would also need to be developed and disseminated in a more participatory and diffused manner than current benchmarking systems, creating a spirit of 'local universality' that is largely missing from the highly centralized and expert-driven models in use today.

Conclusion

As a technical exercise benchmarking may seem objective, although in practice the selection of measurement criteria and the ways in which they are employed in policy are anything but. Dominant forms of benchmarking in the water sector appear to valorize water as a commodity, exclude the public from participation and marginalize alternative worldviews. Deliberately or not, a handful of powerful international actors have designed a monopoly-like form of performance evaluation that emerged from private sector practice which is growing in influence and scope in public water services.

This paper has suggested a possible alternative measurement framework, but developing substitutes will not be easy, particularly if it is seen as a challenge to market norms. It will even be difficult to attract progressive managers and policy makers to the idea of a new benchmarking model, simply because they may not have the time or resources required to do so. The practical and political reality of building an alternative benchmarking framework is an admittedly daunting one.

Nevertheless, there is widespread and growing discontent with water commercialization, and an emerging trend towards rethinking and reclaiming public water, with at least 235 municipalities in 37 countries having remunicipalized their water service from the private sector over the past 15 years, affecting more than 100 million people (Kishimoto et al., 2015; see also Hall et al., 2013; Pigeon et al., 2012). Many more municipalities may revert back to public water in the coming decade, and some of these operators may be looking for new ways of evaluating 'public' water operator performance, offering a timely opportunity to

introduce new methods of assessment alongside new forms of water governance and ownership.

Comparing water operator performance across time and place is inherently problematic, but given the concerns with current practice, and mounting interest in re-making public services, it is a challenge that cannot be ignored.

Acknowledgments

This paper benefitted from comments from a range of academics, practitioners and activists in the water sector including: Madeleine Bélanger Dumontier, Susan Spronk, Marcela Olivera, Mildred Warner, Samir Bensaid, Julie Perkins, Juan Berhau and Klaas Schwartz. I would also like to thank the anonymous referees and David Demeritt, the handling editor at EPA, for their very useful inputs.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Funding for research related to this paper was received from IDRC and from GWOPA.

Notes

1. http://www.iwapublishing.com/template.cfm?name=m2477&utm_source=IWA+Publishing+Mailing+List&utm_campaign=e3de7ca958-GND_9_September_2014&utm_medium=email&utm_term=0_49a7734030-e3de7ca958-90138637 (accessed 9 September 2014).
2. See www.waterbenchmark.org/news/EBCcontributestoMDG-proje.html
3. IBNET – http://www.ib-net.org/en/texts.php?folder_id=78 (accessed 2 August 2014).
4. See www.iwabenchmarking.com/site/documents/
5. See http://www.ib-net.org/en/texts.php?folder_id=91&mat_id=72&L=0&S=0&ss=0 (accessed 7 August 2014).
6. http://www.ib-net.org/en/texts.php?folder_id=78 (accessed 5 August 2014).
7. http://www.ib-net.org/en/texts.php?folder_id=78
8. http://www.iso.org/iso/home/standards_development/standardsdevelopment_gettinginvolved.htm and <http://www.iso.org/iso/home/about.htm>

References

- Abbott M and Cohen B (2009) Productivity and efficiency in the water industry. *Utilities Policy* 17(3–4): 233–244.
- Alegre H, Baptista JM, Cabrera E, et al. (2006) *Performance Indicators for Water Supply Services*. London, UK: IWA publishing.
- Alegre H, Cabrera E and Merkel W (2008) *Current Challenges in Performance Assessment of Water Services*. Water Utility Management International, September, Vol. 3, N. 3, Silver Spring, MD: IWA Publishing, pp. 6–7.
- Alexander N (2005) The roles of the IMF, The World Bank and the WTO in liberalization and privatization of the water services sector. *Citizens' Network on Essential Services*. Available at: www.servicesforall.org (accessed 12 May 2015).

- Aqua Publica Europea (APE) (2014) Aqua Publica Europea position on the European Commission's initiative for benchmarking water quality and services, *APE*, 22 October. Available at: http://www.aquapublica.eu/IMG/pdf/ape_position_on_commission_initiative_for_benchmarking_water_quality_and_services.pdf (accessed 15 August 2015).
- AquaRating (2013) *AquaRating in 13 Slides*. Washington, DC: Inter-American Development Bank. Available at: <http://www.aquarating.org/en/downloads.php> (accessed 6 July 2014).
- AquaRating (2014a) *AquaRating: An Overview*. Washington, DC: Inter-American Development Bank.
- AquaRating (2014b) *AquaRating Commercial Brochure*. Washington, DC: Inter-American Development Bank. Available at: <http://www.aquarating.org/en/downloads.php> (accessed 30 October 2014).
- Baietti A and van Ginneken M (2006) Characteristics of well-performing public water utilities. *Water Supply and Sanitation Working Notes, Note 9*. Washington, DC: World Bank.
- Bayliss K (2014) The financialization of water. *Review of Radical Political Economics* 46(3): 292–307.
- Berg S (2013) *Best Practices in Regulating State-owned and Municipal Water Utilities*. Santiago: Economic Commission for Latin America and the Caribbean.
- Berg S and Corton ML (2007) Water utility benchmarking for managerial and policy decisions: Lessons from developing countries. In: *International Water Association conference on performance assessment of urban infrastructure services*, Valencia, Spain, 12–15 March 2007.
- Breen M and Doyle D (2013) The determinants of privatization: A comparative analysis of developing countries. *Journal of Comparative Policy Analysis: Research and Practice* 15(1): 1–20.
- Boelens R and Vos J (2012) The danger of naturalizing water policy concepts: Water productivity and efficiency discourses from field irrigation to virtual water trade. *Agricultural Water Management* 108: 16–26.
- Bowerman M and Ball A (2000) Great expectations: Benchmarking for best value. *Public Money and Management* 20(2): 21–26.
- Bowerman M, Francis G, Ball A, et al. (2002) The evolution of benchmarking in UK local authorities. *Benchmarking: An International Journal* 9(5): 429–449.
- Braadbaart O (2007) Collaborative benchmarking, transparency and performance: Evidence from The Netherlands water supply industry. *Benchmarking: An International Journal* 14(6): 677–692.
- Cabrera E (2008) Benchmarking in the water industry: A mature practice? *Water Utility Management International* 3: 05–07.
- Cabrera E, Dane P, Haskins S, et al. (2010) *Forget 'Metric' and 'Process', Benchmarking is being Reborn: The New IWA Benchmarking Framework*. Water21. IWA Publishing, December.
- Cabrera E, Dane P, Haskins S, et al. (2011) *Benchmarking Water Services: Guiding Water Utilities to Excellence*. London: IWA Publishing.
- Camp RC (1989) *Benchmarking: The search for industry best practices that lead to superior performance*. Milwaukee, WI: Quality press.
- Clapp J (1998) The privatization of global environmental governance: ISO 14000 and the developing world. *Global Governance* 4: 295–310.
- Corton ML and Berg SV (2009) Benchmarking central American water utilities. *Utilities Policy* 17(3): 267–275.
- Crotty PA (2004) *Selection and Definition of Performance Indicators for Water and Wastewater Utilities*. Denver, CO: American Water Works Association.
- Dahl A and Soss J (2012) *Neoliberalism for the common good?: Public value governance and the downsizing of democracy*. Background Paper, Centre for Integrative leadership, June.
- Danilenko A, van den Berg C, Macheve B, et al. (2014) *The IBNET Water Supply and Sanitation Blue Book 2014: The International Benchmarking Network for Water and Sanitation Utilities Databook, Water and Sanitation Programme*. Washington, DC: World Bank.
- European Benchmarking Cooperation (EBC) (2014) *Methodology Part II: Reference Guide*. Version 5.1. 27 May, mimeo, 206 pp EBC, The Hague.
- Francis G and Holloway J (2007) What have we learned? Themes from the literature on best-practice benchmarking. *International Journal of Management Reviews* 9(3): 171–189.
- Global Water Operators Partnership Alliance (GWOPA). (2009) *Water operators partnerships: Africa Utility Performance Assessment*. Nairobi: GWOPA.

- Hall D, Lobina E and Terhorst P (2013) Re-municipalisation in the early twenty-first century: Water in France and energy in Germany. *International Review of Applied Economics* 27(2): 193–214.
- Harvey D (2000) *Spaces of Hope*. Berkeley: University of California Press.
- Haufler V (2004) Private sector international regimes. In: Bieler A, Higgott R and Underhill G (eds) *Non-state Actors and Authority in the Global System*. New York, NY: Routledge, pp. 121–137.
- International Standards Organization (ISO) (2012) ISO and water: Global solutions to global challenges. Available at: http://www.iso.org/iso/iso_and_water.pdf (accessed 29 May 2014).
- Jackson N and Lund H (2000) *Benchmarking for Higher Education*. New York, NY: Taylor & Francis.
- Jamasb T and Pollitt M (2001) *Benchmarking and regulation of electricity transmission and distribution utilities: Lessons from international experience (No. 0101)*. Faculty of Economics, University of Cambridge.
- Kishimoto S, Lobina E and Petitjean O (eds) (2015) *Our Public Water Future: The Global Experience with Remunicipalisation*. Amsterdam: Transnational Institute, PSIRU, Multinationals Observatory, MSP and EPSU.
- Lefebvre L and Victorisz T (2007) The meaning of social efficiency. *Review of Political Economy* 19(2): 139–164.
- Levy GD and Ronco SL (2012) How benchmarking and higher education came together. *New Directions for Institutional Research* 156(1): 5–13.
- Little JD (1970) Models and managers: The concept of a decision calculus. *Management Science* 16(8): 466–485.
- Mahoney N and Clarke J (2013) Public crises, public futures. *Cultural Studies* 27(6): 933–954.
- McDonald DA and Ruiters GR (eds) (2012) *Alternatives to Privatization: Public Options for Essential Services in the Global South*. New York, NY: Routledge.
- Marques RC and Simões PTF (2010) *Regulation of Water and Wastewater Services: An International Comparison*. London: IWA Publishing.
- Nadvi K and Waltring F (2004) Making sense of global standards. In: Schmitz H (ed.) *Local Enterprises in the Global Economy: Issues of Governance and Upgrading*. Cheltenham: Edward Elgar Publishing, pp. 53–95.
- Osborne D and Gaebler T (1992) *Reinventing Government: How the Entrepreneurial Spirit is Transforming Government*. Reading, MA: Addison Wesley Publications.
- Parena R, Smeets E and Troquet I (2002) *Process Benchmarking in the Water Industry*. London: International Water Association.
- Pidd M (2009) *Tools for Thinking: Modelling in Management Science*. Chichester, UK: Wiley.
- Pidd M (2012) *Measuring the Performance of Public Services: Principles and Practice*. Cambridge: Cambridge University Press.
- Pigeon M, McDonald DA, Hoedeman O, et al. (eds) (2012) *Remunicipalisation: Putting Water Back into Public Hands*. Amsterdam: Transnational Institute.
- Poister TH (2003) *Measuring Performance in Public and Nonprofit Organizations*. San Francisco, CA: Jossey-Bass.
- Pollitt C (2000) How do we know how good public services are. In: Peters G and Savoie DJ (eds) *Governance in the Twenty-First Century: Revitalizing the Public Service*. Kingston: McGill-Queen's University Press, pp. 119–152.
- Parsons LJ (2002) Using stochastic frontier analysis for performance measurement and benchmarking. In: Thomas BF, R Carter Hill, Ivan J, et al. (eds) *Advances in Econometrics*, Vol 16. Bingley, UK: Emerald Group Publishing Limited, pp. 317–350.
- Prakash A and Potoski M (2006) Racing to the bottom? Trade, environmental governance, and ISO 14001. *American Journal of Political Science* 50(2): 350–364.
- Sisson K, Arrowsmith J and Marginson P (2003) All benchmarkers now? Benchmarking and the 'Europeanisation' of industrial relations. *Industrial Relations Journal* 34(1): 15–31.
- Stevenson TH and Barnes FC (2001) Fourteen years of ISO 9000: Impact, criticisms, costs, and benefits. *Business Horizons* 44(3): 45–51.
- Timmermans S and Berg M (1997) Standardization in action: Achieving local universality through medical protocols. *Social Studies of Science* 27(2): 273–305.

- Triantafyllou P (2007) Benchmarking in the public sector: A critical conceptual framework. *Public Administration* 85(3): 829–846.
- Valcik NA, Scruton KE, Murchison SB, et al. (2012) Benchmarking tier-one universities: “keeping up with the Joneses”. *New Directions for Institutional Research* 2012(156): 69–92.
- Van den Berg C and Danilenko A (2011) *The IBNET Water Supply and Sanitation Performance Blue Book: The International Benchmarking Network of Water and Sanitation Utilities Databook*. Washington, CD: World Bank Publications.
- Van der Steen P (2011) *Application of Sustainability Indicators within the framework of Strategic Planning for Integrated Urban Water Management*. SWITCH: The Hague. Available at: http://www.switchurbanwater.eu/outputs/pdfs/W1-1_CALE_MAN_D1.1.7_Indicators_Manual.pdf (accessed 14 June 2014).
- van Helden GJ and Tillema S (2005) In search of a benchmarking theory for the public sector. *Financial Accountability & Management* 21(3): 337–361.
- van Leeuwen CJ, Frijns J, van Wezel A, et al. (2012) City blueprints: 24 indicators to assess the sustainability of the urban water cycle. *Water Resources Management* 26(8): 2177–2197.
- Vos J and Boelens R (2014) Sustainability standards and the water question. *Development and Change* 45(2): 205–230.
- Wait S and Nolte E (2005) Benchmarking health systems: Trends, conceptual issues and future perspectives. *Benchmarking: An International Journal* 12(5): 436–448.
- Water and Sanitation Programme (WSP) (2010) *The State of African Utilities: Performance Assessment and Benchmarking Report*. Washington, DC: Water and Sanitation Programme. Available at: <http://african-utilities-benchmarking.appspot.com/> (accessed 10 August 2014).
- Wolff R (2002) Efficiency: Whose efficiency. *Post-autistic Economics Review*, issue 16, 16 September, article 3, pp. 1–3.
- Zwarteveen MZ and Boelens R (2014) Defining, researching and struggling for water justice: Some conceptual building blocks for research and action. *Water International* 39(2): 143–158.