20 ippr 1988–2008

WWW.IPPR.ORG

# Benchmarking and the Bottom Line: A proposal to improve infrastructure value for money in Britain

## A submission to ippr's Britain's Got Brains competition

by Matti Siemiatycki

October 2008



Institute for Public Policy Research

Challenging ideas – Changing policy

# Contents

## About ippr

The Institute for Public Policy Research (ippr) is the UK's leading progressive think tank, producing cutting-edge research and innovative policy ideas for a just, democratic and sustainable world.

Since 1988, we have been at the forefront of progressive debate and policymaking in the UK. Through our independent research and analysis we define new agendas for change and provide practical solutions to challenges across the full range of public policy issues.

With offices in both London and Newcastle, we ensure our outlook is as broad-based as possible, while our international and migration teams and climate change programme extend our partnerships and influence beyond the UK, giving us a truly world-class reputation for high quality research.

**ippr,** 30-32 Southampton Street, London WC2E 7RA. Tel: +44 (0)20 7470 6100 E: info@ippr.org www.ippr.org. Registered Charity No. 800065

This paper was first published in October 2008. 

## 'Britain's Got Brains'

Britain's Got Brains is a new annual competition run by ippr that aims to identify and promote rigorous, original and exciting policy ideas from postgraduate students and postdoctoral researchers in British universities who may or may not be working in the policy sphere already. The judging panel for this competition includes a wide range of influential politicians, policymakers, journalists, thinkers and academics.

This paper is one of four papers shortlisted by the judges of the 2008 Britain's Got Brains competition. The competition will be run again in 2009. For further details, please visit www.ippr.org/britainsgotbrains.

ippr is grateful to Diageo for its generous financial support of the competition, and to *Prospect Magazine*, the competition's media partner.

Prospect MAGAZINE DIAGEO

## About the author

**Matti Siemiatycki** joined the University of Toronto as an Assistant Professor of Geography and Planning in July 2008. Between 2006 and 2008 he worked as a Research Fellow in the Department of Urban Studies at the University of Glasgow. He holds a PhD in urban planning from the University of British Columbia, Canada. His research focuses on the planning, financing and development of large-scale urban transport infrastructure projects. He has published academic papers on projects in London, Vancouver, Bilbao and Delhi, with recent research examining the procurement of projects through private finance initiatives and the application of benchmarking in project delivery. His forthcoming book examines a century of transportation infrastructure investment in Vancouver. He has also worked as a consultant for Transport Canada and the Greater Vancouver Transportation Authority.

# Executive summary

Infrastructure and public works projects are big business in Britain. In 2007, over £11 billion was spent on constructing new transport, water, waste, school, stadium and hospital facilities. At the same time, in Britain as in countries around the world, new facility costs and delivery times are consistently underestimated and the project benefits overestimated.

For the largest and highest profile infrastructure projects, such as the Scottish Parliament building, the Millennium Dome, Wembley Stadium and the London Underground upgrades, the media has tallied the millions and hundreds of millions of pound overruns caused by rising construction costs and missed delivery deadlines. Beyond these high-profile projects, studies by the National Audit Office show that inaccurate forecasts leading to rising costs are a common feature of government construction projects.

For example, a 2001 review of the construction sector by the National Audit Office found that 73 per cent of government construction projects were completed over the tender price, and 70 per cent were delivered late. More recently, another National Audit Office report of 36 road projects completed by the Highway Agency between 1998 and 2006 concluded that the total cost of delivering the projects was 40 per cent higher than had been initially estimated, with inflation, non-refundable VAT and the spending of contingencies accounting for 92 per cent of the cost increase.

## A new system of benchmarks

This paper develops a system to improve the transparency and accountability of infrastructure project procurement and delivery, which is designed to challenge the underlying causes of overly optimistic forecasts. Specifically, it proposes the development of a system of benchmarks that measure the performance of contractors on past projects, and uses the results to inform the selection of firms for future public contracts. This idea is based on primary and secondary research of similar benchmarking systems that have been developed and implemented in a number of jurisdictions, including Denmark and Singapore. Drawing on the designs of these reference examples it is possible to develop an effective and cost-efficient system for Britain.

In the system proposed here, companies involved in the delivery of large public projects would be ranked based on their past performance in terms of a number of variables that are tailored to the service and type of facilities being provided. For instance:

- Technical consultants would be measured based on the accuracy of their pre-development forecasts.
- Contractors and builders would be evaluated for construction quality and workplace safety, punctuality and cost escalations following the signing of contracts.
- Facility operators would be measured by service availability, cleanliness and maintenance.
- Concessionaires involved in private finance initiatives would be assessed on their success in transferring risks without additional costs to the public sector.
- All government contractors would be evaluated for their customer service, and the ongoing quality of their relationships with the procuring agency and the general public.

The information contained in these benchmarks would then be integrated into the criteria used to select contractors for future government tenders, so that those delivering the highest quality product, on time and on budget would be rewarded by having an increased chance of receiving future work. Over the long term, the widespread availability of benchmarked data comparing performance of individual firms is designed to make corporate reputation and future contract awards more closely based on actual performance, while reducing the significance of perceived success, marketing and public relations. This not only gives public procurement managers greater information upon which to select companies with a strong background in quality performance: it

also gives firms an incentive to drive up the quality of their performance so that they have a better chance of obtaining future contracts. In other words, greater accountability is achieved by linking the short- and long-term interests of the firms involved in planning and delivering Britain's public infrastructure.

The idea of using benchmarking to increase the accountability and performance of project delivery is intuitively quite simple, but has not been systematically applied to date in the British infrastructure sector. The novelty of the system proposed here lies in the flexibility to design the specifications so that they meet the needs of a diverse group of stakeholders, while minimising the cost and time needed to administer the system. The benchmarking database can be administered by a central government agency as in the case of Singapore; by a specially formed non-governmental organisation, as in Denmark, which functions at no cost to government; or by competing private-sector rating agencies that would sell the information to both public and private sector clients. It can be applied to projects delivered through conventional public sector procurement methods or private finance initiatives. And standardisation of the data required to measure a firm's performance can minimise the time needed to participate in the benchmarking exercise, which has been reduced to a reported average of three-and-a-half hours in the Danish system.

In sum, benchmarking the performance of firms delivering public infrastructure in Britain provides a new mechanism to control the forces that contribute to overly optimistic forecasts. This is important since inaccurate forecasts lead to project cost escalations, late delivery and system underperformance that cost the taxpayer dearly, and challenge public confidence in their civic leaders.

# 1. Introduction

When it comes to the delivery of large public sector projects such as highways, public transport facilities, stadiums, hospitals and schools, the media headlines report a familiar story. 'Academy cost overruns hit £48.5m', reported the BBC in 2006. 'Metronet cuts 500 jobs as Tube cost over-run passes £750m', the *Independent* stated in 2007. 'Tram systems too costly and underused', the *Guardian* proclaimed in 2004. 'FA confirms new delays at Wembley', announced the BBC in 2006.

On their own, these media headlines may report on isolated incidents, each with their own set of unique local factors that caused spiralling delivery costs, late delivery and poor operational performance. However, emerging academic research shows a systematic pattern. The most robust study of the performance of large-scale infrastructure projects was conducted by Professor Bent Flyvbjerg and his colleagues at Aalborg University, Denmark in 2003. Focusing on more than 250 large transport projects in 20 countries on five continents, the Aalborg study found that development costs were on average 28 per cent higher than forecasted. Conversely, facility usage forecasts had inaccuracies that averaged between 9 per cent and 39 per cent depending on the type of infrastructure. According to the study, this pattern has not changed over the past 70 years.

In Britain, the major financial costs of inaccurate forecasting on public infrastructure projects are also coming into clearer focus. Construction of new infrastructure and public works in Britain is a significant industry, valued at over £11 billion in 2007 (National Statistics 2008). Beyond the media reports of cost escalations totalling millions and hundreds of millions of pounds on high-profile projects, there is growing evidence that inaccurate forecasting is a common feature of infrastructure project delivery.

A 2001 National Audit Office review of the construction sector reported that 73 per cent of government construction projects were completed over the tender price, and 70 per cent were delivered late. More recently, a 2007 National Audit Office report of 36 road schemes completed by the Highway Agency between 1998 and 2006 concluded that the total cost of delivering the projects was £489 million (40 per cent) above initial estimates, with inflation, non-refundable VAT and the spending of contingencies accounting for 92 per cent of the cost increase.

While forecasting an uncertain future is inherently difficult, technical challenges alone do not explain the persistence of forecast inaccuracies – particularly because the pattern of cost overrun and performance shortfall is highly regular. Rather, contemporary scholarly and professional attention has focused on the way that large-scale public projects are planned, financed, and delivered. It is argued that where there is an absence of long-term accountability mechanisms, some involved parties have a direct incentive to produce or promote unrealistically optimistic forecasts, since they can benefit either financially or politically from the development of a project.

In seeking to minimise the presence of overly optimistic forecasts in the planning of large-scale public projects, a key step is to create institutional cultures that normalise and reward accurate forecasting and construction management, while de-legitimising the practice of producing overly optimistic forecasts. One approach that has been followed in Britain is the introduction of new project delivery methods such as private finance initiatives, which bundle facility design, construction, operation and maintenance into a single concession, and are designed to align project risks and rewards with the party best able to manage them. To date, private finance initiatives make up a relatively small proportion of total public spending, and amidst considerable debate, there is mixed evidence about the degree to which this project delivery mechanism has improved the accuracy of pre-development forecasts.

Alongside both conventional and alternative project delivery methods, an important step in the process to legitimise and reward accurate forecasting is gaining a picture of the firms that consistently deliver their projects on time, on budget, and to a high quality standard, and which firms do not. Intriguingly, this data is not currently available in any organised form for firms involved in the planning, construction, and operation of public infrastructure in Britain. And there are no systematic procedures to aid public sector managers in selecting firms that have a strong record of project delivery.

Within this context, this paper proposes a system that benchmarks the performance of contractors on their past projects, and uses the results to inform the selection of firms for future public contracts. This idea is based on primary and secondary research of similar systems that have been developed and implemented in public sector infrastructure procurement in a number of jurisdictions, including Denmark and Singapore. Drawing on the designs of these reference examples it is possible to develop an effective and cost-efficient system in Britain.

## Structure of the paper

The remainder of this paper sets out the details of the proposal to introduce performance benchmarking into the delivery of large-scale public infrastructure. The introduction of performance benchmarking is a direction that the British government is already undertaking in health, education, and public services, and this proposal fills out the details for expansion in the infrastructure sector.

The first section examines the diverse causes of overly optimistic forecasts in the delivery of large public infrastructure. The second section documents how benchmarking systems in the infrastructure sector have been designed and implemented to date both in Britain and abroad.

Based on this understanding of the diverse set of factors that cause overly optimistic forecasts and a distilling of the lessons learned from the national and international experience with benchmarking, the third section lays out the design of a performance benchmarking and bidder pre-qualification system that can increase project accountability and improve performance in Britain. The final section establishes a strategy to implement a performance benchmarking system in Britain.

# 2. The causes of inaccurate forecasts

Due to the inherent uncertainty of predicting future events, forecasting infrastructure project outcomes is an activity prone to error. Nevertheless, as evidence mounts that project costs and delivery times are consistently underestimated and project benefits overestimated, there is a growing recognition that inaccurate forecasts are not random or only the result of difficulties forecasting uncertain futures.

Rather, a diverse body of literature suggests that the production of overly optimistic forecasts result from the incentive structures created by the social, psychological, political and economic contexts in which projects are planned. Understanding the interactions between these contextual variables is critical to designing systems that can successfully counterbalance the causes of overly optimistic forecasting (Pezzo *et al* 2006).

## Social and psychological causes of inaccurate forecasts

For decades, researchers studying human behaviour have sought to understand the persistent trend of 'planning fallacies', where people tend to underestimate the time and cost that it will take to complete a task. As a type of interaction made between people, forecasts are not only produced through technical processes but also through social processes.

An article in the *Harvard Business Review* by Daniel Lovallo and Nobel laureate Daniel Kahneman (2003) attributes the persistence of overly optimistic forecasts in planning for large projects to a variety of causes. Individuals tend to display overconfidence about their own talents, skills and abilities; they are quick to take personal credit for positive results but tend to attribute the cause of failures to external forces such as inflation or poor weather; and they tend to express perceptions of having greater influence over a situation than may actually be the case, while discounting the role that chance has in achieving either a positive or negative outcome.

The tendency of individuals to accentuate the positive is magnified by forces that occur within organisations. As initial plans for investments are drawn up, future studies become anchored to the technical ideas and cost figures identified in these early plans, and there is often insufficient deviation even as new evidence becomes available. And in a context where organisations have only limited time and resources to pursue new initiatives, large incentives are in place for individuals to accentuate the positives when forecasting the benefits of their own plans. As such, the plans that often get chosen are those imbedded with the most optimistic forecasts (Lovallo and Kahneman 2003).

The social context in which plans are made can also create incentive structures that contribute to the production of biased forecasts. There is considerable evidence that individual and organisational behaviour is most influenced by immediate rather than long-term consequences. In fields where the outcome of a given forecast may not be recognisable for many years, presenting the desirability of an initiative in the short term can outweigh the consequences of making an inaccurate prediction which will not be evident for some time. And when faced with competing pressures to make favourable predictions in the short term and accurate predictions in the long term, the prospect of taking actions to look good in the short term will frequently prevail (Pezzo *et al* 2006).

In response to the social and psychological forces that encourage optimistic forecasting, improving accountability has been proposed as one mechanism to reduce the incentives that introduce bias into forecasts (Sedikides *et al* 2002). To address the challenge of overly optimistic forecasts in particular, Lovallo and Kahneman (2003) suggest that decision-makers use past experience from other projects as benchmarks to compare the accuracy of forecasts for future initiatives.

## Political and economic causes of inaccurate forecasts

Embedded in the social and psychological explanations for overly optimistic forecasts are a range of political and economic incentives that contribute to the production of inaccurate forecasts

during the planning and development of public infrastructure projects. On the one hand, there are many parties that stand to gain either financially or in terms of prestige and image from the delivery of large-scale infrastructure projects. These groups can include politicians, bureaucrats, consultants, lawyers, and contractors involved in the planning and delivery of the project, and property owners and community residents, depending on the type of facility being built.

On the other hand, in contrast to the specific benefits, the allocation of costs when a public infrastructure project does not meet its expectations are not always immediately clear, nor are they often ascribable to the parties that supported or decided to take a specific course of action in the first place. Large-scale infrastructure projects take years to develop and the period over which costs and benefits are intended to accrue can stretch over decades. And in conventional models of public infrastructure delivery where projects are entirely government financed and operated, the majority of the monetary costs for construction overruns, opening delays and revenue shortfalls are often borne by taxpayers, rather than those who planned, approved and supported the project.

The result is that at the same time as there are strong incentives to produce optimistic forecasts that get big projects started, there are traditionally a lack of long-term repercussions for those who are involved with or responsible for the planning and delivery of projects that fail to meet their financial or performance forecasts. For the prominent academics that authored articles with titles such as 'When planners lie with numbers' (Wachs 1988) and 'The lying game' (Flyvbjerg 2003), their research shows that too often the production of inaccurate forecasts is not the result of technical difficulties in predicting the future. Rather, these studies conclude that overly optimistic forecasts frequently result from a systemic pattern of wilful misinformation on the part of project proponents seeking to maximise their individual benefits from an investment initiative. As Bent Flyvbjerg, a leading scholar on the delivery of mega projects, notes, the projects that get built are not

> '…necessarily the best ones, but those projects for which proponents best succeed in conjuring a fantasy world of underestimated costs, overestimated revenues, undervalued environmental impacts and overvalued regional development effects.' (2003: 60)

## Implications from the academic literature

At its core, the production of overly optimistic forecasts is rooted in a series of social, psychological, political and economic incentive structures created by the way that large infrastructure projects are planned and delivered. The literature suggests that accountability mechanisms tailored to the specific challenge of delivering public infrastructure provide a potential avenue to create new incentive structures that limit the benefit from making overly optimistic forecasts.

The accountability mechanisms that work best are ones that link the short- and long-term interests of those producing project forecasts, so that the benefit of striving to minimise overly optimistic forecasts translates into a stronger reputation for those making more accurate estimates. Future rewards to those making accurate forecasts can encourage forecasters to implement internal procedures that counteract many of the factors that cause over-optimism, and ultimately lead to projects that better meet their initial expectations.

# 3. International application of benchmarking: changing the incentive structure

League tables have become popular in a variety of sectors of both public and private service delivery in Britain, including further education, health, financial services and consumer products. Recently, in Britain and abroad, public sector agencies have also started to compile databases of corporate performance for firms involved in the planning, construction and delivery of major infrastructure projects. According to a Dutch study, as of 2004 there were approximately 40 different benchmarking systems being used in the construction and infrastructure development sector worldwide (Bakens *et al* 2004). Briefly reviewing the design of a sample of these systems in Britain and around the world provides insights into the current state of practice, and the factors that contribute to effective performance benchmarking in the infrastructure sector.

## England: The Highways Agency

In Britain to date, public-sector-led pre-qualification systems in the infrastructure sector are typically based on the financial strength of the company, the structures it has in place to manage and deliver projects and its capabilities to carry out projects based on existing workload, rather than detailed evaluations of past performance of the projects that the company has delivered (FHWA 2005).

The Highways Agency in England is a leading example of an organisation that has developed a benchmarking system for companies, and has used the data to pre-qualify bidders for future tenders beginning in 2002 (FHWA 2005). Participation in the Highways Agency's pre-qualification system is voluntary, and requires a self assessment of a company's capabilities and management systems, which is then verified by external auditors (Highways Agency 2006).

There are three key challenges identified with the system to date:

- First, despite employing a 288 point rating scale, all of the firms scored in a fairly narrow range, and thus the evaluation has not been fine enough to identify major differences in performance between the 33 companies that participated.
- Second, the system primarily evaluates firms based on their inputs, rather than the quality of their outputs.
- Finally, it has been reported that producing a benchmark entry can cost a firm up to half a million pounds in time and resources, which some contractors claim is too expensive. (Pearman 2007)

## Denmark: An innovative approach to benchmarking in the construction sector

A system implemented in the Danish construction industry in 2005 provides an example of the possibilities of an enhanced benchmarking system that focuses on project outputs. Denmark is a particularly useful point of comparison since it has a similar political and legal structure to the UK.

Faced with a historical record of poor performance in the Danish construction sector, in 2001 members of the construction industry partnered with the Danish Government to set up the Benchmarking Centre for the Danish Construction Industry (BCE). The BCE was fully funded by the private sector as a commercial foundation, with a mandate to develop and disseminate information as part of a benchmarking system. Nearly half of its revenue is generated through the sale of project evaluation materials to industry sources such as contractors and sub-contractors; the remainder of its funding comes from industry and private foundation contributions.

In 2003, the Danish government passed legislation making it mandatory that all firms bidding for state government contracts over the equivalent of £500,000, excluding infrastructure projects, must demonstrate their occupational safety, punctuality, quality capabilities, profitability and customer satisfaction based on their performance on past projects. To meet these objectives, the BCE developed a grade book system, where for a sample of government contracts that a company performed, their

performance was evaluated by the public sector client based on a series of 14 key performance indicators. The accuracy of the data supplied by the companies is verified by the BCE, and the grade book is then submitted to a government procurement agency along with the company's tender proposal for each new job. This provided a new level of data on which decisions about contract allocation can be made (Mortensen and Lind Kristensen 2006).

According to a report by the BCE, the collaborative approach was critical to obtaining support for the benchmarking initiative from the construction industry:

> 'Introduction of compulsory benchmarking constitutes an administrative burden on companies and is not necessarily a popular measure. The reason the initiative has succeeded is largely due to the fact that BCE was formed by the organisations in the construction sector, which have seats on the Directors Board and have been backing the system actively.' (Mortensen and Lind Kristensen 2006: 5)

Through close public and private sector collaborations in the design of the benchmarking system, the typical time that it takes for a company to carry out their requirements for the benchmarking of a single project is reported to be 3.5 hours, and ongoing efforts are being taken to reduce this time further.

In Denmark, the implementation of a robust benchmarking framework makes it possible to compare companies based on their inputs and output performance, and the data has been used to carry out extensive studies of performance within the industry and develop best practices. A key limitation of the Danish benchmarking system is that it was primarily designed to pre-qualify bidders above a minimum standard, and to date there has not been widespread action to formally integrate the benchmarks into the final process of selecting the winning bidder. This means that there is not necessarily a strong ongoing incentive for firms to strive to improve their performance over time above a minimum standard needed to pre-qualify for government contracts. A second limitation is that resulting from a separation in the government departments responsible for construction and infrastructure when the plan was drafted, the benchmarking system does not apply to infrastructure projects, even though these are typically the most costly and highest profile types of public projects.

## Singapore

The construction quality assessment system set up for public sector residential building in Singapore (CONQUAS) is an interesting example of a system that provides a direct and ongoing financial incentive to companies that are assessed as delivering high-quality outputs on their previous projects. The CONQUAS system is particularly designed to measure the architectural and construction quality of a firm's output, and other service aspects.

For every point over a score of 65 that a company receives on their quality assessment, they are awarded a tendering advantage of 0.2 per cent up to a maximum of 5 per cent or the equivalent of £1.9 million (whichever is less) on future contracts. This means that for instance if a company receives a score of 75, they will be awarded a contract if their bid is within 2 per cent of the price of their nearest competitor, although they will be paid at the original bid price (Tam *et al* 2000).

Despite being only a small premium, Kam and Tang (1997) found that the incentive programme has been effective. As a result of the financial incentive built into the tendering process that rewards ongoing quality development, the average ranking and number of companies on the premium list has increased steadily, reflecting a rise in the overall quality of construction on public sector projects in Singapore.

## Lessons from the international experience

Overall, the international evidence shows that when appropriately designed and integrated into the procurement process, bidder pre-qualification systems based on the benchmarking of corporate performance can effectively provide companies with incentives to deliver high-quality outputs. An important insight from the international experience is that the benchmarking of corporate

performance on past projects alone is not enough to ensure improved quality on future projects. Rather, as in the system implemented in Singapore, formalised structures that integrate the benchmarking results into the final stages of future government procurement processes provide the strongest incentives to companies to deliver high-quality projects.

In reviewing the merits of pre-qualification procedures, a report by the United States Federal Highway Administration (2005) noted that such systems have the potential to limit competition between firms by introducing non-cost variables into the evaluation process. It was argued that ultimately, adding evaluations of firms' past performance to the criteria used to assess current bids could lead to higher priced projects in the long term, since competition may be reduced during the tendering process. Despite this challenge, the report recommended that American jurisdictions undertake the development of pre-qualification systems in the transport sector, since such systems can improve the alignment of short- and long-term interests of the private and public sector to better achieve mutually-agreed objectives.

Supporting their recommendation, the report states:

> 'When discussing construction management issues of quality assurance/quality control, contract change processes, environmental monitoring, etc., they [public sector managers in countries that have developed performance benchmarking systems] frequently state that contractors have incentives to perform these practices well because they know it will affect their ability to participate in future work, either directly through past performance rating or indirectly through an assessment of their qualifications and capabilities.' Federal Highway Administration (2005: 16)

In addition to changing the terms on which competition occurs from lowest cost to best value, the introduction of performance benchmarking has the potential to create more transparent and accountable procurement processes that are to the benefit of both public sector clients and high-performing companies. The widespread availability of benchmarked data comparing performance of individual firms has the potential to make corporate reputation and future contract awards more closely based on actual performance, while reducing the importance of industry perceptions, marketing and public relations. This benefits firms that have a strong record delivering high-quality outputs, and provides an incentive for all companies to drive up their performance levels. Moreover, when the benchmarked data is formally integrated into the selection process, it removes a key data blind spot that challenges public sector decision-makers, who may not have complete information about the actual performance of firms on previous contracts that were carried out for other government bodies or in other jurisdiction.

# 4. Detailed design of a proposed benchmarking system in Britain

Having examined the diverse causes of overly optimistic forecasts in the delivery of large-scale infrastructure, and explored how performance benchmarking and bidder pre-qualification has been applied internationally, this section sets out the design of a system that can be implemented in the British infrastructure sector. The system is purposely designed in a flexible manner so that it is applicable to different types of projects, as well as projects procured through both conventional models of project delivery and private finance initiatives.

## Who to benchmark?

At present, the benchmarking systems that have been implemented in the infrastructure sector typically focus on the project management and construction phases of the delivery process. However, in many countries including Britain, an increasing share of project scoping, planning, design, testing,

risk management and operation functions for the delivery of large public infrastructure are being conducted by private companies. This is the case whether projects are delivered and funded through private finance initiatives or by more conventional public sector models of delivery. In the transport sector, for instance, a comparative review of road procurement practices in North America and Europe found high levels of outsourcing in most jurisdictions, with nearly all aspects of project delivery carried out by the private sector in some countries (FHWA 2005).

In this context, where private sector firms are playing an increasing roll in all facets of project definition and delivery, it is necessary that performance benchmarking and pre-qualification systems are in place to maximise performance at all stages of the procurement process. Leagues should be constructed to compare the performance of firms that deliver similar types of infrastructure (roads, bridges, rail lines, waste and water facilities, schools, hospitals), and provide similar types of services (financial evaluations, performance forecasts, project management, construction, system operation). A special category of benchmark should also be created to compare how successfully concessionaires involved in private finance initiatives transfer and control the risks that they have taken on.

Public infrastructure projects vary considerably with respect to size, timeframe of delivery, and technical and political complexity. This creates a potential for unfair performance comparison between firms working on different types of projects.

Overcoming this challenge requires an approach that seeks as best as possible to compare the work of firms on like projects. To this end, benchmarks should compare the work of companies within defined project types and project price ranges. For instance, one benchmark would be for companies producing forecasts or constructing bridge projects up to £500 million, another for bridge projects over £500 million. In some cases it may be useful to subcategorise even further, for example to differentiate between urban railway projects that require tunnelling (a major source of cost overruns and project delays) and those where the entire project is at grade.

Varying scales and complexities within a single league can also be handled by assigning a complexity weighting to each project evaluated as part of a company's benchmark calculation. This factor would seek to account for variations in political and technical challenges that may impact a company's performance on a project.

## What to benchmark? A focus on inputs and outputs

As performance benchmarking and pre-qualification systems are applied to a wider range of sectors involved in the procurement of public infrastructure, performance benchmarking systems should continue to compare companies based on simple input data, such as company size, safety systems in place and employee training programmes available. Detailed information covering issues such as management systems, profitability per project, detailed delivery methods and corporate strategies should not be collected. This information is both costly to collect and commercially sensitive, with its public release potentially conferring unfair advantages between competing firms.

In addition to the collection of simple input data, there is also a need for greater emphasis on data measuring product outputs. In particular, as has been implemented in the Danish system of benchmarking for the construction industry, benchmarks should be produced comparing the differences between forecasted and actual costs, punctuality of product completion, price, safety, and quality of delivery outputs and customer satisfaction.

Table 4.1 outlines the types of information that should be collected and benchmarked to compare the quality of work performed by companies working at each stage of public project delivery. As the benchmarking system is developed in greater detail to meet the needs of the British infrastructure sector, the evaluation criteria outlined in Table 4.1 should be translated into specific key performance indicators that can be used to measure and compare the performance of each firm.

All of the data being collected as part of the proposed benchmarking system is information that is already contained within public documents or information that can be obtained through freedom of information laws. For this reason, there are no legal barriers to releasing the data publicly.

**Table 4.1. Prospective information types for project benchmarking**

| Project procurement stage | Private sector function | Cost | Punctuality | Output quality |
|---|---|---|---|---|
| Planning | Adviser to Government on project cost | Is the report produced for the agreed cost?<br><br>What is the average hourly charge-out rate for consultants on the project?<br><br>How does the total price of producing the report compare to studies on similar projects? | Is the report delivered on time? | How accurate were cost projections as compared to the cost of the actual winning bid, and the final construction price? |
| Planning | Adviser to Government on system usage* | Is the report produced for the agreed cost?<br><br>What is the average hourly charge-out rate for the government advisers? | Is the report delivered on time? | How accurate were usage projections compared with actual system use at defined intervals following opening? |
| Delivery | Project management and construction | Is there a discrepancy between the contractually established and actual cost of the facility?<br><br>Was the Government required to make extra payments above those contractually agreed upon in order to complete project construction? | Is there a discrepancy between the projected and actual opening of the facility? | Are health and safety and labour laws observed during construction? How many serious injuries per person hour were recorded on site?<br><br>Does the project construction meet the agreed specifications with respect to material standards, architecture and design, and craftsmanship? |
| Operations and maintenance | Project operation and maintenance | Is the Government required to make extra payments in order to maintain levels of service? | Are major maintenance tasks carried out at agreed intervals? | Does the service meet contractually agreed performance standards?<br><br>Are health and safety and labour laws observed during operation and maintenance? |
| Private finance initiative:<br><br>Design-build-finance-operate | Design, deliver, operate and finance the project through a single concessionary contract | Did the cost of the concessionaires bid change between being selected as the best proposal and the signing of the final contract?<br><br>Was the Government required to make extra payments to the concessionaire in order to complete the project once it was started?<br><br>Did early investors gain large profits through major debt restructurings or stock sales?<br><br>How did the total price of the project compare to that of similar projects? | Is there a discrepancy between the projected and actual opening date of the facility? | Does the service meet contractually agreed performance standards?<br><br>Are health and safety and labour laws observed during operation and maintenance?<br><br>What is the accident and injury frequency for workers on the facility?<br><br>Were disputes resolved through the contractually mandated resolution mechanisms, or was the legal system used? |

Note: *The definition of usage varies depending on the type of infrastructure being developed. For instance, usage would be defined as the daily number of patrons carried on a new public transit system, how many kilowatts/hour of electricity are produced by a power plant, or the litres of sewage treated by a new plant.

## Data collection, cost and accuracy

Any effort to embed greater data collection and monitoring into the infrastructure project delivery process will come with considerable implementation and operation costs. A key hurdle for any benchmarking system is to balance the necessity for accuracy and comprehensiveness against the time requirements and costs involved with participation for both the Government and the companies.

There are two strategies to mitigate the costs of implementing and operating an expanded corporate benchmarking and pre-qualification system. First, collaboration should be sought between the public and private sector to encourage corporate submission of data in a format that can be easily integrated into the benchmark, rather than requiring auditors to mine the data from detailed technical studies. To this end, data sharing on project performance can be built directly into the planning, project delivery process, and evaluation stages of procurement, making it less costly to monitor and measure corporate performance. In Denmark, the data submission process has been streamlined so that it reportedly takes around 3.5 hours in total.

Second, as is the case in benchmarking systems in Denmark and Singapore, data is collected for a sample of projects rather than all projects carried out by a company, unless there is specific evidence suggesting the need for a more in-depth audit. As the benchmarking system is implemented in Britain, stakeholder consultations and pilot studies should be carried out to measure whether benchmarking based on a sample of project outcomes provide sufficient levels of accuracy.

## Structuring the benchmarking system

The firms involved in the infrastructure sector in Britain typically work on projects across the country. To accurately reflect their performance and provide a complete picture of regional variations in the quality of infrastructure delivery, the benchmarking system should be 'scalable', with the data organised in a hierarchy so that trends can be observed at the firm, project type, city, regional and national scale. The data should also be structured so that comparisons can be made between projects from different sectors such as waste, water, transportation, health and education, and also between different types of project finance and delivery mechanisms.

## Administering and managing the benchmarking system

There are three models that could be used to develop, manage and administer the benchmarking system: directly through a government department or crown corporation; through a not-for-profit foundation; or by private companies. Each model has strengths and weaknesses, which are compared in Table 4.2 (next page). The most effective model for application in Britain should be selected following extensive consultations to determine the requirements of industry, government and community stakeholders.

## Dissemination and uses of the benchmarked data

There are multiple users for the benchmarked data. Public sector procurement managers can use the information on past performance in order to select contractors with the best record of delivering services on time and on budget. As noted above, benchmarking is most effective at improving future performance when the data is formally integrated into the tender evaluation and selection process. The model employed in Singapore, where firms are graded based on past performance and given a small financial advantage for future projects tenders, provides the clearest method of providing long-term incentives for companies to deliver high-quality outputs.

The data can also be used as part of the research and development process in the infrastructure sector. In Denmark, researchers have used the benchmarked data from the construction sector to identify regional or inter-sectoral variations in the success of project delivery, and to study the impact of input variables such as company size, experience and contract type on project performance. These initial findings can then be used as the basis for more in-depth studies and to guide public policy.

**Table 4.2. Approaches to administering performance benchmarking**

| | Private sector companies | Non-profit organisation | Public sector department or Crown corporation |
|---|---|---|---|
| International example | Company credit ratings (e.g. Standard and Poors; Moody's) | Danish benchmarking in the construction sector | Ontario pre-qualification for transport sector; Singapore housing benchmarks |
| Structure | Private companies design benchmarks and collect data for infrastructure sector. Public and private clients purchase data as required for each project. Firms also pay to have data benchmarked. | Non-profit organisation formed with board of directors from government and industry. Design, administer and disseminate data. | Government department or agency designs and administers benchmarks for all infrastructure sectors. Disseminates data publicly and to project planners. |
| Strengths | 1. Minimal public bureaucracy or initial cost to government.<br>2. Competition between firms can lead to design of best measures and reduced data costs. | 1. Forum for collaboration between firms and government.<br>2. Can be non-partisan. | 1. Data available to government project planners.<br>2. Designed specifically to meet needs of public procurement agents. |
| Weaknesses | 1. Benchmarks may be designed to maximise revenues from consumers, by favouring large companies with many projects.<br>2. Public release of data may be limited to protect resale value. | 1. Organisation may lack authority or power to shape the benchmark system when faced with strong public and private agencies.<br>2. May require government funding. | 1. System may be seen by industry as too closely aligned with government, and therefore they may object to participating.<br>2. Significant cost to taxpayers.<br>3. Conflicting role if government is watchdog and purchaser of public services. |

There is also a private market for the benchmarked data. As private financing of public infrastructure becomes more prevalent through the application of private finance initiatives, banks, pension funds and insurance companies are increasingly putting large sums of money at risk in the infrastructure sector. With returns often based on the degree to which construction costs are controlled and patronage and revenue forecasts are achieved, rigorous benchmarking of past performance by forecasters, designers, engineers, general contractors and facility operators can serve as an important addition to investors' existing due diligence process.

Finally, the benchmarked data should be available to the general public. The collection of commercially sensitive data has been purposely avoided so that the release of the benchmarks does not reveal trade secrets or adversely affect competition between firms. In Britain, public release of benchmark data is in line with the reporting of benchmarks from other sectors, such as the rankings of university departments and hospitals. Access to this information will enable citizens to better scrutinise the performance of the firms that are delivering projects in their communities, and meaningfully engage in project planning processes as they are ongoing.

## Monitoring performance

There may be unintended consequences associated with the expansion of corporate benchmarking systems to include a wider range of project outputs, such as delivery punctuality and meeting budgets. In particular, in order to obtain high benchmark scores and therefore gain advantage in obtaining future contracts, it is possible that firms may rush their jobs thereby creating unsafe work environments, reduce product quality in order to meet their cost obligations, or undertake aggressive measures to reduce labour costs.

To ensure that benchmarks based on project cost and timely delivery do not sacrifice employee safety, product quality, or minimum pay standards, ongoing monitoring of safety standards on job sites and post-implementation quality reviews are a critical component of the benchmarking exercise, as is the inclusion of key performance indicators that cover these issues.

## Legal liability and dispute resolution

The effectiveness of an expanded pre-qualification system based on corporate benchmarking will only be effective if the results accurately reflect the performance of the companies measured. At certain stages of the project delivery process such as construction and during operation and maintenance, information can be collected as part of ongoing quantitative project performance monitoring exercises, as has been carried out in the Singaporean benchmarking system. In this case, it is possible to link outcomes such as substandard quality of materials or workmanship, or unavailable services to the firm responsible for that task.

However, accurately benchmarking performance outcomes becomes more complicated, and potentially controversial, in situations where poor outcomes are not merely the result of failures by individual companies, but are introduced by either the commissioning agency itself or by errors that are the result of outputs produced by a group of firms. This is particularly the case for work carried out in the early phases of project planning with respect to forecasting future costs, completion times, and usage levels, although it is also a challenge during construction and operation.

The extent to which public sector agencies are the source of cost escalations varies depending on the project, but a study of cost increases in public projects delivered through private-public partnerships in Scotland found that 65 per cent of all cost increases were exclusively the result of public sector causes (Cambridge Economic Policy Associates 2005). Moreover in many cases, contractual ambiguity results in contention about who is to blame for major cost overruns or delivery delays, and legal proceedings have often been launched to remedy disputes.

In projects where multiple private sector partners are involved as government advisers or in project delivery roles, forecast inaccuracies and poor corporate performance can also result when an error by one firm is embedded in the subsequent work of others. For instance, financial forecasts are often based on estimates of future infrastructure usage levels, which may have been produced by a different firm from the one that conducts the financial modelling. In this example, errors in the usage predictions will contribute to errors in the results of the financial modelling.

To protect themselves against errors made by other companies or contributing parties, planning consultants and project advisers typically include wide-ranging disclaimers as part of their reports. As an example, when providing a government agency in Vancouver with technical advice on delivering a new rail project as a private-public partnership, a large international consulting firm included the following 'IMPORTANT NOTICE' on the first page of the study:

> 'In preparing this study, Macquarie has relied on information that is publicly available and/or information provided by various other parties. This information has not been independently verified by Macquarie. Macquarie believes to the best of its knowledge that the statements contained in this study are accurate. Macquarie makes no representation or warranty (whether express or implied) as to the accuracy or completeness of this study or its contents.' (Macquarie, 2003: 1)

There are three approaches that can be employed to address the challenge of accuracy in the production of the corporate performance benchmarks.

First, prior to being produced, it is important that there are transparency and explicit instructions about how contentious issues such as cost escalations and substandard work by external firms will be considered as part of the benchmarking system. In the benchmarking systems employed in Denmark and the Canadian province of Ontario, for instance, firms conducting work as part of a partnership or a consortium are held responsible for the outputs of all of their members, while general contractors are responsible for the output of firms carrying out subcontracted work (MTO Qualification Committee 2006).

Second, appeal and dispute resolution mechanisms should be clearly established for situations where there are disagreements over responsibility and benchmarked scores for specific projects.

Third, the benchmarked results should be seen as separate from the issue of legal liability over cost overruns and poor project outcomes. To this end, rather than serving as a post-priori mechanism to assign blame, the benchmarked results tied to a pre-qualification system are instead designed to provide firms with a long-term interest in minimising the forces that cause overly optimistic pre-development forecasts.

The development of a league table system that ranks forecasts against actual outcomes creates an incentive system for companies to question the logic underpinning the stated or unstated assumptions of their public sector clients, which are often a source of error. Also, rather than indemnify themselves against external sources of error as is currently common in the industry (as exemplified in the Vancouver example), firms would have a greater incentive to verify the accuracy or quality of the inputs into their own products, and select project partners that have a strong record of good performance. This would provide another check on the types of biases that can occur early in a project and become magnified over the course of the planning and delivery process.

# 5. Conclusions and implementation strategy

Around the world, more than 40 public sector agencies use pre-qualification systems based on performance benchmarking as part of their infrastructure procurement processes. To date there is evidence that such systems have contributed to improvements in the quality of project delivery.

In order for the next generation of bidder pre-qualification systems to be effective, it is necessary that they are not perceived by companies to be punitive of poor past performance. Rather, they must be seen as a legitimate tool that supports and rewards companies that deliver high-quality outputs, punctually and on budget. Research by Howard Davis from the Warwick Business School and colleagues, published by the Joseph Rowntree Foundation, suggests that systems of external inspection to maximise the quality of public service provisions derive their legitimacy through three forces: the use of credible inspectors; transparent criteria of service evaluation; and providing information and incentives that contribute to the improvement of services (Davis *et al* 2004). As the next generation of corporate performance benchmarking and pre-qualification systems are developed for the infrastructure sector in Britain, extensive industry, civil service, stakeholder and public consultation are necessary, to identify the precise conditions for their successful application.

With this in mind, below is the outline of a strategy to implement a performance benchmarking system in the British infrastructure sector, which draws on the collaborative and iterative process that was used to develop and implement benchmarking in the Danish construction industry.

**1. Meet with members from key stakeholder groups in the public sector, private sector, and community, to determine the applications and parameters that would make a corporate performance monitoring system most useful and effective.**

- The purpose of these initial meetings is to inform the ultimate design of the procurement monitoring system, by obtaining information on how it may be used.
- In the public sector, interviews with a sample of senior procurement agents from a variety of government departments will provide insight into the types of performance that should be monitored, and the ways that the league tables of corporate performance should be integrated into existing decision-making frameworks. Information will also be generated on the way that the output could be most effectively disseminated.
- In the private sector, interviews will be conducted with a sample of corporate managers in the consulting, engineering, project management, construction, operations and infrastructure finance sectors. The purpose of these interviews is to integrate the opinions of the infrastructure industry into the design of the system on a number of key issues, including: the types of variables that should be monitored in each sector, the ways that the benchmarked information is integrated into future tendering processes and publicly disseminated, and how the system should be administered.

**2. Compile a pilot database that contains information on the performance of individual companies to meet their forecasted projections on completed projects.**

- Initially, a pilot database may be created based on a single sector in a single industry. For instance, building on recent studies of road construction costs by the National Audit Office, an initial league table may be produced comparing the accuracy of consultants' estimated and actual costs of building new road projects in Britain.
- This database will build on data that is already publicly available in the UK Partnerships project database, National Audit Office and HM Treasury department studies, as well as data derived from technical planning documents.
- The specific variables used to compare performance will be determined based on the results of the interviews with private and public sector stakeholders, as well as a review of the data available from various sources.

**3. Develop quantitative and qualitative measures in order to analyse whether the league table results reveal significant differences in the performance of individual companies in a given industry.**

- These measures must be easily understood, while accurately reflecting the relative performance of different companies in a given sector or industry. In circumstances where a number of performance variables are being used, relative weightings will be assigned to each variable in order to develop a composite ranking.
- Since infrastructure projects are complex, involving multiple partners and external factors, a common framework is necessary that fairly accounts for the range of circumstances that influence outcomes, which are often beyond the control of a given company. This must be done sector-by-sector and industry-by-industry, in order to account for the contextually specific factors that influence performance.
- As part of developing the measures, tests should be undertaken in order to determine whether a sample of projects from each company provides an accurate measure of their performance, or whether a full census of projects is necessary to gauge performance.

**4. Develop guidelines to integrate the benchmarked results into public sector decision-making frameworks.**

- A consistent set of guidelines should be developed so that the league table results of past performance inform future decisions, without being such a significant element of the selection process that they necessarily preclude any company that meets a minimum ranking from obtaining tenders. In this sense, a system is required so that the league tables can be used in an informative, rather than punitive manner.
- At this stage, a new round of stakeholder meetings should be conducted in order to provide information about the interim results of the study, and obtain feedback on the league tables as developed. Fine tuning to the league table variables and their recommended applications can be carried out to take into consideration public feedback.

**5. Integrate the developed league table into a pilot procurement.**

- As part of the pilot, a monitoring programme should be developed to examine the effectiveness of using the league tables of past corporate performance as part of the procurement criteria. The monitoring programme should examine whether there was any adverse impact on the number of bids submitted or the competitiveness of the tenders, as well as the ways that public sector procurement managers used the information as a component of their decision-making process.

**6. If successful, develop league tables for a wider array of sectors and industries, and create specific guidelines to integrate their usage into the project procurement process.**

# References


Altshuler A and Luberoff D (2003) *Mega-Projects: The Changing Politics of Urban Public Investment* Washington, DC: Brookings Institution

Ahadzi M and Bowles G (2004) 'Public–private partnerships and contract negotiations: an empirical study' *Construction Management and Economics* 22: 967–978

Bakens W, Vries O and Courtney R (2005) *International Review of Benchmarking in Construction* Netherlands: PSI Bouw

Cambridge Economic Policy Associates (2005) *Public Private Partnerships in Scotland: An Evaluation of Performance* Edinburgh: Scottish Executive

Davis H, Downe J and Martin SJ (2004) *The Changing Role of Audit Commission inspection of Local Government?* York: Joseph Rowntree Foundation

Debande O (2002) 'Private Financing of Transport Infrastructure: An Assessment of the UK Experience' *Journal of Transport Economics and Policy* 36(3): 355-387

Federal Highway Administration (FHWA) (2005) *Construction Management Practices in Canada and Europe*. Available at: http://international.fhwa.dot.gov/construction_mgmt/index.htm

Flyvbjerg B (2004) *Procedures for Dealing with Optimism Bias in Transport Planning: Guidance Document* London: The British Department for Transport

Flyvbjerg B, Holm S and Buhl SL (2004) 'What Causes Cost Overrun in Transport Infrastructure Projects?' *Transport Reviews* 24(1): 3-18

Flyvbjerg B, Bruzelius N and Rothengatter W (2003) *Mega Projects and Risk: An Anatomy of Ambition* London: Cambridge University Press

Flyvbjerg B (2003) 'The Lying Game', *Eurobusiness* June: 60-62

Hall P (1982) *Great Planning Disasters* London: Penguin Books

Highway Agency (2006) CAT3 Supplier Information, available at: www.highways.gov.uk/business/documents/CAT3_PH1_Suppliers.pdf

Institute for Public Policy Research (2004) 'Private sector backs thinktank call for greater openness in PFI deals', press release, available at: www.ippr.org.uk/pressreleases/?id=792

Kam CW and Tang SL (1997) 'Development and implementation of quality assurance in public construction and works in Singapore and Hong Kong' *International Journal of Quality & Reliability Management* 14(9): 909-928

Lovallo D and Kahneman D (2003) 'Delusions of Success: How optimism undermines executives' decisions' *Harvard Business Review* 81(7): 56-63

Macquarie (2001) *PPP Review of RAV Rapid Transit Project: Part One Final Report* Vancouver: Macquarie

Mortensen S and Lind Kristensen E (2006) *Benchmarking Danish Construction Copenhagen: The Benchmark Centre for Danish Construction,* available at: www.byggeevaluering.dk/db/files/bec_uk_2006.pdf

National Audit Office (2007a) *Estimating and Monitoring the costs of building roads in England* London: The Stationery Office

National Audit Office (2007b) *Improving the PFI tendering process* London: The Stationery Office

National Audit Office (2001) *Modernizing Construction* London: The Stationery Office

National Statistics (2008) 'Output and Employment in the Construction Industry', web page, available at: www.statistics.gov.uk/STATBASE/Product.asp?vlnk=725

Pearman R (2007) 'Balfour Beatty tops Highways Agency's CAT score league', *Contract Journal,* 14 March

Pezzo SP, Pezzo MV and Stone ER (2006) 'The social implications of planning: How public predictions bias future plans' *Journal of Experimental Social Psychology* 42 (2): 221-227

Pickrell D (1992) 'A Desire Named Streetcar: Fantasy and Fact in Rail Transit Planning' *Journal of the American Planning Association* 58(2): 158–176

Public Works Finance (2006) 'Survey of Private Public Partnerships' *Public Works Finance* 209: 1-8

Sedikides C, Herbst KC, Hardin DP, and Dardis GJ (2002) 'Accountability as a deterrent to self-enhancement: The search for mechanisms' *Journal of Personality and Social Psychology* 83(3): 592-605

Serco Institute (2008) Institute Resources, available at www.serco.com/instituteresource/

Siemiatycki M (2006) 'Message in a Metro – Building Urban Rail Infrastructure and Image in Delhi India' *International Journal of Urban and Regional Research* 30:2: 277-292

Tam CM, Deng ZM, Zeng SX and Ho CS (2000) 'Performance assessment scoring system of public housing construction for quality improvement in Hong Kong' *International Journal of Quality & Reliability Management* 17(4/5): 467-478

Throgmorton J (1991) 'Planning as a Rhetorical Activity' *Journal of the American Planning Association* 59(3): 334-346

Wachs M (1988) *When planners lie with numbers: An exploration of data, analysis and planning ethics* Los Angeles: University of California