Despite the significant volume of studies on public sector performance measurement, a paucity of empirical research describes in detail the systems and processes used at different levels of government to measure and manage performance. This study focuses on the experience of Public Service Agreements in the public sector in England. In particular, the impact of a centralized, performance measurement-driven approach on public service delivery is analyzed using case studies in a health care and a police organization. Despite efforts to introduce a “golden thread” to link different levels of the public sector hierarchy, in both cases, there was relatively low consistency in terms of performance indicators, targets, and priorities. Significant implications are evident for the design and role of performance targets and indicators, for the possibility to align frameworks at different levels of the public sector, and for the importance of feedback loops in measurement systems.

The challenges of measuring performance—in both the public and private sectors—have been discussed for decades (Behn 2003; Carter, Day, and Klein 1992; Hood 2006; Johnson and Kaplan 1987; Neely 1999; Pollitt and Bouckaert 2004; Ridgway 1956; Ridley and Simon 1938). Governments are demonstrating growing interest in the measurement of performance in the public sector, as are the press and media, and officials are using performance targets and league tables in order to push through modernization programs and demonstrate that value for taxpayers’ money is being delivered.

Observers of these changes coined the phrase “New Public Management” to describe them (Hood 1991, 1995). In many advanced economies, such as those of the Anglo-Saxon countries, Scandinavia, and the Netherlands, public services have come under increasing pressure to improve their efficiency and effectiveness, reduce their demands on taxpayers, but maintain the volume and quality of services supplied to the public (Brignall and Modell 2000). Recent empirical studies have found that performance measurement is growing in both state and local governments in the United States (Breul and Kamensky 2008; Wichowsky and Moynihan 2008). Indeed, state governments “have enthusiastically embraced the idea of managing for results” (Moynihan 2006, 77), adapting policy ideas transferred from the United Kingdom, New Zealand, and Australia. However, significant improvements in performance and results-based accountability have not been fully achieved, and research is still needed to identify the key determinants of successful design, implementation, and use of performance measurement systems (Sanger 2008).

In the United Kingdom, the government has put considerable emphasis on the provision of public services since the early 1990s. As a Royal Statistical Society report states, “A striking feature of U.K. public services in the 1990s was the rise of performance monitoring, which records, analyses and publishes data in order to give the public a better idea of how government policies change the public services and to improve their effectiveness” (2005, 2). Along the same lines, Hood, James, and Scott remarked that “the U.K.’s New Public Management era cannot be adequately understood without reference to this growth of public-sector regulation” (2000, 301).

Over the last decade, an important innovation in the performance management regime in the public sector in England has been the introduction of Public Service Agreements (PSAs). In essence, PSAs are explicit agreements, targets, and indicators that are established between the finance ministry (known as Her Majesty’s Treasury) and individual government departments, which subsequently are cascaded throughout the public sector in an effort to ensure delivery alignment. This paper looks at a specific period—the early 2000s—during which PSAs played a very important role in the government’s widely publicized public service delivery strategy. Using empirical data from two sectors—the police and the health services—this paper assesses the effectiveness of Public Service Agreements and the overall performance measurement regime on the public sector in England.
This research does not aim to provide conclusive evidence about the impact of performance measurement systems on performance. Rather, it considers the effects of a centralized, performance-measurement-driven approach on public service delivery. Although research undertaken in the field of performance measurement supports the use of performance measurement systems as a means to align strategy and action, the case of the PSA system in England tells a more nuanced story. In particular, it is interesting to note that, notwithstanding the substantial investment of resources and emphasis placed on such system, this performance measurement regime failed to establish a pervasive and consistent approach that could link the different elements of the “delivery chain.”

Observers used the phrase “golden thread” to highlight one of the aims of the [Public Service Agreement (PSA)] system—that is, the attempt to ensure the overall coherence of the “chain” of government-set targets from the national to local levels. Golden threads are present if objectives, targets, and indicators are consistent throughout the different levels from central government to service delivery. This research explores the extent to which golden threads were established as a result of the introduction of Public Service Agreements in England.

In the next sections, the English performance measurement framework is introduced, followed by the research design and the analysis of data collected in two areas of the public sector—a health care organization and a police force. Subsequently, the findings are discussed and implications for the wider research and practitioner communities presented.

The English Performance Measurement Framework

At the heart of the English performance measurement framework is the Public Service Agreement regime. PSAs were first introduced in 1998 and form a central plank in the contract between the Treasury and government departments. In the United Kingdom, spending reviews are conducted every two years. During these, the Treasury and government departments negotiate departmental priorities, deliverables, and budgets. The agreed priorities and deliverables are encapsulated in the PSAs, which are published upon completion of the spending review process. To assess delivery, each PSA consists of a number of objectives, targets, and an accompanying technical note, which explains how progress toward each target will be measured. The objectives within departments’ PSAs are designed to encapsulate the full scope of the departments’ responsibilities. However, PSA targets focus on departments’ key priorities—for example, areas where a national aspiration is considered appropriate in order to drive forward a step-change in performance, or where it is appropriate to set a national “standard.”

From a Treasury perspective, the PSA objectives, targets, and accompanying technical notes are the major elements of the public sector performance measurement framework. Once agreed, the Treasury recognizes that departments will seek to translate the PSA objectives, targets, and technical notes into more detailed statements, but the Treasury plays no formal part in this cascade process. The number of PSA targets has progressively decreased, from 250 in 1998, to 160 in 2000, down to 130 in 2002, and 110 in 2004. In the 2007 spending review, this figure was reduced even further, with 30 PSA targets (disaggregated into approximately 180 performance measures) being agreed upon. This article uses data from 2002—a period in which Public Service Agreements were arguably at the pinnacle of their history, as the engine of a centralized, performance-measurement-driven system.

The Golden Thread

Over the past two decades, scholars have considered the use of performance measurement systems an effective means for implementing organizational strategies and ensuring alignment between strategy and action in private firms (Chenhall 2005; Dixon, Nanni, and Vollmann 1990; Ittner, Larcker, and Randall 2003). By following a consistent cascading process, it is claimed, performance measurement systems can enable the alignment of departments and business units with the corporate center, thus leading to higher financial performance (Kaplan and Norton 2006).

This article argues that a somewhat similar perspective was adopted in the public sector in England. Indeed, the performance measurement regime introduced in the late 1990s and early 2000s aimed at ensuring consistency between national policy and local delivery. Observers used the phrase “golden thread” to highlight one of the aims of the PSA system—that is, the attempt to ensure the overall coherence of the “chain” of government-set targets from the national to local levels. Golden threads are present if objectives, targets, and indicators are consistent throughout the different levels from central government to service delivery (Audit Commission 2002).

Research Design

This research explores the extent to which golden threads were established as a result of the introduction of Public Service Agreements in England. To do so, the authors mapped the performance measurement framework that was in operation in part of the English public sector in the early 2000s. Case studies were chosen as the most appropriate research strategy (see Carpenter and Ferez 1992; Kaplan 1983). Operationally, two exploratory case studies (Yin 1994) were carried out in two different areas of the public sector—health care and police. These two cases are particularly significant, as the introduction of PSA targets proved controversial in both areas, and conflicting evidence has been found as to the benefits and shortcomings of this centralized, performance-measurement-driven approach (Bevan and Hood 2006; Butterfield, Edwards, and Woodall 2004; Collier 2001; Kelman and Friedman 2007; Propper et al. 2008; Royal Statistical Society 2005).

From an empirical point of view, these case studies enabled the researchers to illustrate in one case the linkages between the performance measurement systems used by the Treasury, Department of Health, and Healthcare Commission and the indicators used by a primary care trust that is directly controlled by the National Health Service and, to a large degree, by the central government; in the other case, the relations between the performance measurement system developed by the Treasury, Home Office, and one of the country's...
Police forces and health care trusts show substantial differences on a number of levels: the services they deliver, the environments they work in, and the diversity and number of skills they have to mobilize to deliver those services (cf. Carter, Day, and Klein 1992). However, given the great influence of the central government on how performance was measured, there was a high degree of similarity in the ways in which performance targets and indicators were designed across levels in both services. Therefore, this research looks at the two cases separately in order to identify specific issues, but also compares them to search for common aspects related to performance measurement and the concept of the golden thread.

The data collection methods used include document analysis and, to a lesser extent, unstructured interviews with key informants. The information gathered through documents and interviews was analyzed consistently, and the NVivo software was utilized for the coding. This analysis also involved an examination of the language used and the identification of possible differences in the emphasis placed on specific aspects. Given the relevance of public accountability and disclosure of performance data to the public in the U.K. government discourse, publicly available documents were used as a primary source of information.

The analysis is presented in the next sections and consists of two main phases. The researchers started by assessing the internal consistency of the elements of the organizational performance measurement systems (i.e., whether the targets and indicators within an organization were consistent). Second, the consistency of the performance measurement system designed by the different organizations within the two sectors is examined. In the conclusive sections, patterns emerging from the analysis of the data are presented and reflections are made regarding their relevance for an international audience.

**Health Care**

**Departmental Perspective**

In the English health care system, various bodies are involved in the design and development of performance targets and indicators. First, the Treasury, through the Department of Health, details the aims, objectives, and targets of the whole sector, according to the previously formulated spending review. Under the 2002 spending review, the Department of Health agreed to an overarching aim, three PSA objectives (including one value-for-money target), 12 associated targets, and 32 resulting indicators (figure 1).

Second, the Healthcare Commission—the independent inspection body for both the National Health Service and independent health care—sets specific targets and indicators for the different trusts (acute, ambulance, mental health, and primary care). Third, before each primary care trust applies or adapts these indicators to its own particular situation, several other bodies intervene and influence the way they are structured. In essence, a primary care trust is responsible for improving the health of the local population, providing community and primary care services, and commissioning hospital and other specialist services. Primary care trusts control about 80 percent of the total National Health Service budget.

The analysis of published material enabled the identification of four organizations or frameworks that played a major role in the measurement and management of performance: the priorities and planning framework; the personal social services performance assessment framework; best value; and the strategic health authorities. Finally, the last level of this hierarchical structure consists of individual primary care trusts, which formulate their objectives, targets, and indicators.

In the sections that follow, the linkages between the performance measurement systems developed by the Department of Health and Healthcare Commission and how they relate to a primary care trust—the West Suffolk Primary Care Trust—are examined.

**Department of Health and Healthcare Commission**

Looking at the Department of Health’s documents, there clearly was a high level of consistency among the different elements of the system—that is, targets were consistent with objectives, and indicators were good metrics for the targets to which they referred. The examination of the documents pertaining to the Healthcare Commission provides a very different picture. In this case, the...
system consisted of key targets and balanced scorecard indicators. However, the two were by no means related. This contradicts the general assumption in the performance measurement literature that targets and indicators should be clearly connected. Furthermore, it is not clear what it was meant by “balanced scorecard,” as the framework originally developed by Kaplan and Norton (1992, 1996) refers to a totally different model than the one proposed by the Healthcare Commission. It is interesting to notice that the description of the key targets also included the specification of the indicators that made them operational, and that the balanced scorecard indicators were structured in the same way as the key targets. Confusion resulted over what was considered a target and what was considered an indicator.

Further coding of documents and analysis of discourse show how key targets and balanced scorecard indicators often referred to different issues; in the two areas in which there were similarities (inpatient/outpatient waiting and staff surveys), there were differences between the sources of data (information comes from different databases) and the time scales of the indicators. Even in areas of interest such as access, where two targets and two balanced scorecard indicators mapped in each category, it is possible to see how targets and indicators differed. The key targets, in this case, referred to “the percentage of patients able to be offered a routine appointment to see a GP (or a PCP) within one working day.” The first balanced scorecard indicator, on the contrary, was “constructed by aggregating responses from the survey questions which relate to access and waiting.” The second concerned “single telephone access through NHS Direct to GP out-of-hours care,” which still referred to access, but by measuring a very specific aspect, not included in the key targets. Therefore, it can be argued that the main difference between targets and indicators seemed to be in the use of technical terms. It appears that the Healthcare Commission could have renamed their key targets as “general targets and indicators,” and the balanced scorecard indicators as “specific targets and indicators,” without significant loss of form or function.

**Analysis of the Different Perspectives**

After examining the internal structure of the performance measurement systems developed separately by the Department of Health and the Healthcare Commission, the consistency between the two systems will now be evaluated. The Department of Health sets the overarching aims, objectives, targets, and indicators for all public health care organizations. The Healthcare Commission independently sets targets and indicators for the acute, ambulance, mental health, and primary care trusts. The analysis, through the grouping and comparison of concepts (i.e., the nodes of the coding structure), shows that several common issues existed, including access to professionals; patient satisfaction; inpatient, outpatient, and accident and emergency room waiting times; and some aspects regarding certain categories of patients (e.g., children and older people) and specific treatments or technical targets (e.g., drug users, cancer patients, and cardiac patients) (see table 1).

Furthermore, to better assess the degree of similarity between the targets and indicators described in the two documents, several terms and phrases (48 in total, obtained through coding) were interrogated. These searches confirmed that some issues were dealt with in both documents, but they also emphasized several differences:

- The Department of Health targets and indicators and the Healthcare Commission key targets show some similarities, but correspondence with the balanced scorecard indicators is limited. This is partly attributable to the level of detail that the documents divulged—the balanced scorecard indicators often involved technical terms, while the Department of Health document and the Healthcare Commission key targets did not.
- Although there are targets that belonged to similar areas, the way in which performance indicators were designed was different in most of the cases (e.g., outpatient and inpatient waiting, suicide prevention, etc.).
- The Healthcare Commission documents show a strong focus on both service delivery (patient satisfaction) and staff satisfaction, while the Department of Health document considered only service delivery (the word “staff” is never used, while in the Healthcare Commission document, it is used 37 times). Furthermore, while appraisals and surveys were core concepts in the Healthcare Commission document, they were not even mentioned by the Department of Health.
- The formulation of the Department of Health document was strongly influenced by the New Public Management lexicon, in contrast to the Healthcare Commission document. Keywords such as “accountability,” “effectiveness,” “efficiency,” and “value for money” could be found in the Department of Health text, but not in the Healthcare Commission text.
- Some words or phrases that could be expected to appear in such documents are totally absent in the Department of Health document (e.g., equity, human resource, ethnic, financial, information, safety, staff, smoking, etc.).
- Comparing the two documents, it seems that although the objectives, targets, and indicators developed by the Department of Health had the aim of informing citizens, the Healthcare Commission mixed similar elements with metrics that would be used to assess the way in which various trusts operated. This means that external indicators concerning service delivery, for instance, were sometimes mixed with internal indicators regarding imposed mechanisms (e.g., audits correctly carried out). This generated confusion about the goals and role of the performance measurement system.

The comparison of the Department of Health and Healthcare Commission documents emphasizes a substantial degree of incongruence and, certainly, an insufficient explicit linkage between the two. The Department of Health never mentioned the Healthcare Commission, whereas the Healthcare Commission referred just seven times to the Department of Health and only twice to PSAs. In this respect, a more explicit connection might have been expected, as well as an explanation of the existence of certain targets and indicators. Finally, looking at the main elements of the two documents, it is possible to see how the Department of Health targets correspond only partially to the Healthcare Commission ones. More specifically, the analysis shows how just four of the PSA targets were totally covered, six were partially covered, and two were not covered at all.

Following the same procedure used in the case of the police force, the researchers compared the documents produced by the Department of Health and Healthcare Commission with those produced by the West Suffolk Primary Care Trust. This analysis investigated the differences between the targets and indicators used.
by the West Suffolk Primary Care Trust and the ones set by the Department of Health and Healthcare Commission. In fact, the degree of correspondence was very low. The comparison shows that just about one-fifth of the indicators used by the primary care trust were present in the two other documents. A further 15 percent were used in one, but not the other, and in more than 40 percent of cases, it was not possible to find any correspondence. A major reason is that, as with the case of the police, significant primary care trust funding comes from joint initiatives with local authorities and other delivery agencies, and each of these initiatives tends to have its own monitoring and evaluation framework, with associated indicators.

**Police Force**

**Departmental Perspective**

Under the 2002 spending review, the Home Office agreed to an overarching aim of seven PSA objectives (including one value-for-money target) with the Treasury, with 10 associated targets and 37 indicators. Once the PSA objectives, targets, and technical notes have been agreed with the Treasury, individual government departments determine how they will break down these high-level objectives into specific initiatives and actions. There is no mandated process for doing this, and hence each department has discretion over how it translates the PSA objectives and associated targets into language that is meaningful and appropriate. This section focuses on the approach adopted by one major department—the Home Office—and explores how this department translates its PSA targets into an appropriate performance measurement framework for the English police force. Examining the 37 indicators that the Home Office negotiated with the Treasury, it is possible to see how the first nine and the last one applied specifically to the police force. Therefore, these are the indicators considered here.

In 2002, the Home Office released the first National Policing Plan, covering the period 2003 to 2006. In essence, the Home Office's

---

**Table 1 Comparison of Department of Health and Healthcare Commission Groups**

<table>
<thead>
<tr>
<th>Groups</th>
<th>Department of Health, Targets and Indicators</th>
<th>Healthcare Commission, Key Targets (and relative indicators) and Balanced Scorecard Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access</td>
<td>Access to professionals (access to a GP; access to a PCP)</td>
<td>Access to a GP (appointment with a GP); Access to a PCP (appointment with a PCP); PCT patient survey—access; single telephone access</td>
</tr>
<tr>
<td>Financial</td>
<td>Value for money (improvement in value for money; cost-efficiency indicator; service effectiveness indicator; service effectiveness indicator)</td>
<td>Financial management (relative unplanned financial support; other financial indicators; surplus or deficit; unplanned financial support; variance from financial plan)</td>
</tr>
<tr>
<td>Internal</td>
<td>Hospital appointments (inpatients admitted; outpatients booked)</td>
<td>Delayed transfers of care; health equity audit; NHS dentistry; PCT commissioning; suicide audit</td>
</tr>
<tr>
<td>Patient satisfaction</td>
<td>Hospital appointments (inpatients admitted; outpatients booked)</td>
<td>PCT patient survey—information; PCT patient survey—relations; PCT patient survey—facilities; PCT patient survey—safe care</td>
</tr>
<tr>
<td>Particular groups of patients</td>
<td>Access to professionals (access to a GP; access to a PCP)</td>
<td>Access to a GP (appointment with a GP); Access to a PCP (appointment with a PCP); PCT patient survey—access; single telephone access</td>
</tr>
<tr>
<td></td>
<td>Financial management (relative unplanned financial support; other financial indicators; surplus or deficit; unplanned financial support; variance from financial plan)</td>
<td>Financial management (relative unplanned financial support; other financial indicators; surplus or deficit; unplanned financial support; variance from financial plan)</td>
</tr>
<tr>
<td></td>
<td>Internal</td>
<td>Internal</td>
</tr>
<tr>
<td></td>
<td>Patient satisfaction</td>
<td>Patient satisfaction</td>
</tr>
<tr>
<td></td>
<td>Particular groups of patients</td>
<td>Particular groups of patients</td>
</tr>
<tr>
<td></td>
<td>Staff</td>
<td>Staff</td>
</tr>
<tr>
<td></td>
<td>Treatments and technical targets/indicators</td>
<td>Treatments and technical targets/indicators</td>
</tr>
<tr>
<td></td>
<td>Waiting</td>
<td>Waiting</td>
</tr>
</tbody>
</table>

---

Performance Measurement in the Public Sector in England 595
view in 2002 was that the PSAs would translate into Best Value Performance Indicators, which could be used to track performance at the local level. In addition, the 2002 plan made reference to the Policing Performance Assessment Framework (PPAF), which would have developed “a set of national standards on performance to ensure a consistent, first class service” (U.K. Home Office 2002, 32). The PPAF consisted of six domains—citizen focus, helping the public, reducing crime, investigating crime, promoting public safety, and resource use. At the time considered in this study, the English police forces were assessed on the basis of an interim framework, consisting of 49 separate indicators, documented in the National Policing Plan. These 49 indicators clustered around three broad categories. Six of them are identified as relating to PSA 1:

- Reduce crime and the fear of crime; improve performance overall, including by reducing the gap between the highest crime in Crime and Disorder Reduction Partnerships areas and the best comparable areas; and reduce:
  - Vehicle crime by 30 percent from 1998–99 to 2004
  - Domestic burglary by 25 percent from 1998–99 to 2005
  - Robbery in the 10 Street Crime Initiative areas by 14 percent from 1999–2000 to 2005
  - Maintain this level

Another 12 indicators are classified as relating to PSA 2:

- Improve the performance of all police forces, and significantly reduce the performance gap between the best and worst performing forces; and significantly increase the proportion of time spent on frontline duties.

The remaining 31 indicators are derived from the Best Value Regime and mapped onto the six dimensions of the PPAF framework.

**Local Perspective**

For the police, the local level is the county or region. The researchers chose to focus on one of these regions—Cambridgeshire. In Cambridgeshire, two further plans were developed—the Police Authority Strategic Plan and the Local Delivery Plan. The first of these plans covered the time period 2003 to 2005, while the second covered 12 months.

From the analysis of documents, it emerged that, while, in theory, the local-level indicators should cascade neatly from PSAs to PPAF to Local Delivery Plans, in practice the Local Delivery Plans had also to take account of (1) National Policing Plans, (2) Local Policing Plans, (3) Best Value Key Performance Indicators, (4) Constabulary Service Charters, (5) Efficiency Plans, (6) Community Safety Strategies, and (7) Financing and Resourcing Plans.

This multitude of influences resulted in 78 performance indicators used to track performance at the local level (see figure 2). Importantly, while at the Treasury and Home Office levels, the PSA and associated indicators were most widely discussed, at the local level, another performance measurement framework (Best Value) appeared to be influential. This was probably because, while central government saw the Best Value Regime as outdated, the regime was enshrined in the statute books through the Local Government Act 1999 and hence was still fairly important at the local level. Furthermore, it is important to notice how in some police Basic

---

**Figure 2 Performance Indicators Used to Track Performance at the Local Level in a Police Force**

<table>
<thead>
<tr>
<th>Aim of the Home Office: Build a safe, just, and tolerant society</th>
<th>Treasury perspective: 6 objectives, one value for money target, and 37 separate indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSAs and Technical Notes</td>
<td></td>
</tr>
<tr>
<td>National Policing Plan: 2004–2007</td>
<td>Department perspective: 6 domains and 49 interim indicators</td>
</tr>
<tr>
<td>Police Performance Assessment Framework</td>
<td></td>
</tr>
<tr>
<td>Police Authority Strategic Plan: 2003–2005</td>
<td>Local perspective: 78 separate indicators</td>
</tr>
<tr>
<td>Local Delivery Plan</td>
<td></td>
</tr>
<tr>
<td>National Crime Reduction Strategy</td>
<td></td>
</tr>
<tr>
<td>Best Value Performance Plan</td>
<td></td>
</tr>
<tr>
<td>Constabulary’s Service Charter</td>
<td></td>
</tr>
<tr>
<td>Local Priority Plan</td>
<td></td>
</tr>
<tr>
<td>Efficiency Plan</td>
<td></td>
</tr>
<tr>
<td>Community Safety Strategies</td>
<td></td>
</tr>
<tr>
<td>Finance and Resourcing Plan</td>
<td></td>
</tr>
</tbody>
</table>

---
Command Units, up to 50 percent of funding came from “projects and special initiatives” rather than “base budgets” (through Crime and Disorder Reduction Partnerships, special Home Office schemes, etc.). Each of these separate short-term funding streams tended to have its own monitoring and evaluation frameworks, generating separate performance indicators.

**Analysis of the Different Perspectives**

Some interesting and important issues emerge from the analysis of documents. The first point to note is the rate of growth of indicators through the system. Through the PSA process, the Treasury and Home Office agreed to seven PSAs and 37 separate indicators, of which three PSAs and 10 indicators were relevant to the police. In the Home Office’s interim framework, these three PSAs and 10 indicators were converted into 49 separate indicators. At the local level, the 49 indicators identified in the Home Office’s interim framework were converted into 78 separate indicators. Therefore, the cascade process effectively managed to convert the three PSAs into 78 separate performance indicators, giving an average of 26 indicators per PSA. If every department had followed the same process, then the 130 PSA targets defined in 2002 would have translated into approximately 3,380 indicators at the local level. Therefore, it is important to question why this happens and whether the resources required to track this number of indicators are justified. In the case of the police, it appears that the answer to this question is that the local level performance indicators were derived from the Best Value review framework introduced by the Labour government when it first came to power. Indeed, it could be argued that there was a disconnection between the Treasury’s PSA Regime and the Best Value Regime.

Second, the documents show that there seemed not to have been sufficient reflection on the relationships between the indicators. Moreover, little explicit sense of priority among the variety of targets and indicators had been established both at the local and departmental levels. Strategy maps (Eccles and Pyburn 1992; Kaplan and Norton 2004; Neely, Adams, and Kennerley 2002) could be used to overcome these problems, as they are designed to help people identify which are the critical indicators and to understand—or at least endeavor to articulate—the relationships between the different indicators. Potentially, this approach might prove valuable to those in the public sector designing measurement systems as they seek to rationalize the number of indicators they use.

The third point involves a more detailed analysis of the relationships between the indicators. For each of the 78 indicators, the researcher explored whether it could be mapped directly back onto (1) the original PSAs, (2) the 10 PSA indicators, and (3) the 49 interim indicators identified in the Home Office’s interim framework. This analysis demonstrated that two-thirds of the local-level indicators could be linked directly to the 10 PSA indicators through the 49 interim indicators. However, if the two PSA indicators—7 and 8—are removed, then this figure drops to just one-third of the local indicators. PSA indicators 7 and 8 referred to “performance in the six domains of the Policing Performance Assessment Framework,” which in itself was a very broad-ranging framework. The breadth of these two PSAs is illustrated by the fact that about 75 percent of the local-level indicators can be mapped onto them. This in itself calls into question their validity.

It is interesting to carry this analysis further: of the 10 PSA indicators, there are 160 mappings—that is, 160 separate occasions in which local-level indicators map onto specific PSA indicators. Of these 160 mappings, four PSA indicators—1, 3, 7, and 8—account for more than 90 percent of them. This suggests that the indicators used at the local level were heavily skewed toward two broad categories of PSA indicators—those that relate to the overall level of crime and those that relate to the PPAF framework. It also calls into question why the remaining six PSA indicators had been included at the same level in the framework, or at least questions whether much progress would have been made against them.

The same analysis can be applied to the interim indicators. In this case, there were 138 mappings, split across 42 indicators. For seven interim indicators, there was no corresponding local-level indicator, whereas 10 interim indicators accounted for more than half of all mappings, suggesting a strong bias of attention toward certain three specific areas: violent crime, public satisfaction, and fear of crime.

**Discussion**

Comparing the results obtained in the two case studies, a number of common areas of concern emerge. First, in both the police and health service, the level of consistency in terms of indicators, and consequent targets and priorities, was relatively low across the hierarchy that constitutes the public sector. Despite central government and regulatory emphasis on performance measurement and substantial use of resources, the practice and theory of the “golden thread” in the public sector in England was not as widespread as one might have thought.

Second, looking at the linkages between performance targets and indicators at different levels in both sectors, the analysis shows a substantial growth of indicators through the system and a high degree of confusion at the local level regarding priorities. Third, in none of the cases was it possible to see evidence of adequate reflection made on the relationships between performance indicators, and rarely was it feasible to identify connections between targets and indicators.

Following this analysis, it is clear that the conflicting interests and different rationales existing at different levels of the public sector generated confusion locally. Organizations such as the Cambridgeshire police force and the West Suffolk Primary Care Trust needed access to resources if they were to deliver service. They had to adopt measurement targets and indicators that could be used in the political negotiation process. And yet this is where the system broke down. For it was not always clear at the local level who held the power, and hence access to finance, at the central level. In this particular performance measurement regime, organizations at the local level were forced to respond to the requests of the PSA framework imposed by the Treasury, but also to frameworks imposed by others—for instance, the Best Value regime. Moreover, in the case of the police force, mechanisms of local accountability implied that the organization had to reconcile national priorities with local needs, which often were different, and pressure from the media. The Cambridgeshire police force is a particularly good example, as performance management and formal reporting became quite confused when national politicians and the media got involved in a high-profile murder case during the period studied in this article. It was effectively these competing demands that caused much of the
Implications for the Policy and Research Communities

The debate over the impact of Public Service Agreements on public service delivery is still open (see Hood 2007; Institute for Government 2008; Smith 2007), and it is not the aim of this article to provide conclusive evidence in this sense. However, this research has considerable implications for scholars, policy makers, and practitioners who are interested in the measurement and management of public services. Each of these roles has important implications for the design of the performance measurement systems and their implementation. It is clear that the performance measurement systems in both organizations considered suffered from trying to fulfill too many of these roles at the same time. Indeed, as performance targets and indicators are developed as part of a framework, their results could only generate information that is meaningful in the context of that framework (Mari 2007). Therefore, using the same measure for different purposes could lead in many cases to confused messages and counterproductive approaches to performance measurement and management. Although citizens are mainly concerned about improvements in service delivery, the findings of this study suggest that this role was the least enabled.

Second, certain tools or practices could be adopted to improve the overall process of design and use of performance targets and indicators. In the case of strategy maps, private sector companies have made much progress in their application, and it is believed that public sector organizations could benefit from them, too, as this is a general tool that is neither context nor sector specific.

Third, the evidence presented in this paper contributes to the debate on the alignment of performance measurement systems within and among organizations. Although it is recognized that performance measurement systems could play an important role in driving service delivery, implementing strategy, and influencing behavior and action, it is questionable whether the establishment of a pervasive “golden thread” would be either feasible or desirable in the public sector. Research on alignment has mostly been conducted in private sector contexts, with the organization as the typical unit of analysis, and with a focus on the links between corporate center and business units. Even leaving issues related to the negotiation of targets and budgets aside (see, e.g., Lindblom 1977; Wildavsky 1975), this paper shows the difficulty, if not the impossibility, of creating a fully aligned system enforced through performance targets and indicators. Indeed, this approach, coupled with quite a centralized, top-down approach, is likely to clash with the variety of purposes and roles that a performance measurement system could and should fulfill, and not deliver on its promise of public service improvement.

Finally, several major points emerge from this study. First, while investigating measurement in the public sector, the identification of the role of the measurement system becomes a primary concern. It could be argued that targets and indicators are designed to control local delivery, improve accountability, compare organizations’ performance, and influence behavior and action in order to improve public services. Each of these roles has important implications for the design of the performance measurement systems and their implementation. It is clear that the performance measurement systems in both organizations considered suffered from trying to fulfill too many of these roles at the same time. Indeed, as performance targets and indicators are developed as part of a framework, their results could only generate information that is meaningful in the context of that framework (Mari 2007). Therefore, using the same measure for different purposes could lead in many cases to confused messages and counterproductive approaches to performance measurement and management. Although citizens are mainly concerned about improvements in service delivery, the findings of this study suggest that this role was the least enabled.

This research demonstrates that, despite substantial resources employed and emphasis put on the “delivery chain” by central government—in terms of financial rewards, reputation, and autonomy for local organizations—there was no coherent and clear set of priorities for the English public services, which aligned with a “golden thread” running from central government to local delivery organizations.
of objectives throughout the different levels that constitute the public sector could be regarded as a positive result, this could not be achieved through an almost purely top-down approach. In fact, the existence of feedback loops from the bottom up is a necessary component of an effective measurement system (Simons 1995). This feedback mechanism would provide essential information about whether the objectives, goals, and targets are practical, feasible, and desirable. Also, it would increase the commitment and contribution of local level organizations and, ultimately, the participation of citizens in local service delivery.

Notes

1. Within the United Kingdom, different performance measurement regimes have been implemented. Therefore, because the empirical analysis presented in this paper is based on the study of public sector organizations in England, the research findings relate only to England and not to the whole of the United Kingdom. It should also be remarked that, despite the progressive devolution of powers to Scotland and Wales since 1999, considerable central control is still a dominant feature of English government (Hood 2006).

2. These are the overall level of crime, as measured by the British Crime Survey; fear of crime; crime rate in the highest Crime and Disorder Reduction Partnerships; vehicle crime (British Crime Survey); domestic burglary (British Crime Survey); robbery (police-recorded crime); performance in six domains of the Policing Performance Assessment Framework: the performance gap between the best and worst performing forces; time spent on frontline duties; and report on progress toward police efficiency gains.

References


---

**Movement and Change:**

**“Transitions in Public Service”**

*Rio Hotel*

Las Vegas, Nevada

September 30 – October 2, 2010

The 2010 National Association of Schools of Public Affairs and Administration (NASPAA) annual conference will present panels, research papers, roundtables, and workshops. The conference tracks are: transitions in our universities, agency transitions, and political transitions.

For more information, go to: