Performance Management in Practice: A Comparative Study of Executive Agencies

Christopher Pollitt
Erasmus University Rotterdam

ABSTRACT

This article reports a study of performance management practices in four functions across four European Union member states (Finland, the Netherlands, Sweden, and the United Kingdom). The focus is on how and to what extent performance indicators influenced the top management of the agencies concerned and the degree to which performance data were used by ministries as steering instruments. The research uses a historical institutionalist perspective combined with a model that identifies primary task characteristics as a source of significant variation. Thus the design explores both the influence of task characteristics (through contrasts among the four different functions) and embedded national system characteristics (through contrasts among the four countries). I show that both primary task characteristics and national system characteristics had some of the theoretically predicted effects on the management regimes. Equally, however, certain general tendencies embraced all countries and most functions. These include, first, the incremental growth of more sophisticated performance indicator systems and, second, the feebleness of ministries in developing performance-based strategic steering.

Over the past two decades the introduction of performance management and the creation of executive agencies have been two of the most widespread international trends in public management (De Bruijn 2002; Holzer and Yang 2004; Ingraham, Joyce, and Donahue 2003; Joustie 2001; Mol and Kruijf 2004; Office of Public Services Reform 2002; Organization for Economic Cooperation and Development [OECD] 2002; Pollitt and Bouckaert 2004; Pollitt and Talbot 2004; Pollitt, Talbot, Caulfield, and Smullen 2004; Van Thiel 2004; Yesilkagit 2004). While these two (linked) types of development have been “translated,” in different ways in different countries and sectors, there can be little doubt of the pervasiveness of the reform rhetoric featuring one, or often both, of these “solutions.” As yet, however, we are rather short of comparative empirical research into the actual practice of “agencification” and its relationship to performance management.

This article is a single-authored derivative of the EUROPAIR project, which was a team effort involving Janice Caulfield, Amanda Smullen, Colin Talbot, and myself. Adrian O’Reilly and Karen Bathgate also assisted in the early stages. EUROPAIR was supported by a grant from the U.K. Economic and Social Research Council and by the Centre for Public Management, Erasmus University Rotterdam. Address correspondence to the corresponding author at Pollitt@fsw.eur.nl.

doi:10.1093/jopart/mui045
© The Author 2005. Published by Oxford University Press. All rights reserved.
For permissions, please e-mail: journals.permissions@oupjournals.org.
This article takes a step toward remedying that shortage. It reports selected findings from an in-depth study of a matched set of agencies in four countries: Finland, the Netherlands, Sweden, and the United Kingdom. One aim was to get behind the surface rhetoric of performance management in order to examine actual practice (Pollitt 2002). The idea was to be able to compare the performance management arrangements for roughly the same primary tasks in all four countries. The tasks are forestry, meteorology, social security, and prisons. Most of these were conducted by agencies, although there were a few cases where other types of autonomous public bodies carried out the task (see table 1). A general treatment of the findings can be found in Pollitt et al. 2004.

The research was focused mainly on the respective influences of the national systems (the culture and institutional pattern in each country) and tasks (would there be strong similarities among, say, forestry bodies in the four countries and differences with, say, meteorological offices?). Both sets of factors have been cited in the scholarly literature as affecting the ways in which generic ideas are received and “translated” in particular jurisdictions. In this case, therefore, we wanted to know how far national systems/cultures and task characteristics influenced the way the generic ideas of performance management and agencification were translated into each national and functional context.

AIMS

The first purpose of this article is to report comparative research findings. The main research questions are as follows:

1. How was “performance” conceptualized and measured in the agencies studied? In other words, what were the extent and sophistication of performance measurement?

2. To what extent were performance measurement systems integrated with other major management systems—specifically, financial management, human resources management, and planning? In other words, what was the extent of performance management?

3. How far, by whom, and for what were the performance data used? In other words, what was the extent of performance steering, especially by the supervising ministries?

4. How do the answers to the above three questions vary with country and with task/function?

A second purpose is to connect these findings to the relevant theoretical literature and, in particular, to assess the usefulness of combining a historical institutionalist approach with a focus on task-specific characteristics.

THEORETICAL BACKGROUND

Although the research was primarily empirical and descriptive, it could not, of course, pretend to be theory free. Our approach was shaped by a combination of two theoretical traditions. First, we were persuaded by the many works that conclude that national cultures (e.g., Hofstede 2001; Lynn, Heinrich, and Hill 2001; Mouritzen and Svara 2002) and/or national institutional patterns (e.g., Christensen and Lægreid 2001; Gains 2004; Kickert 2000; Lijphart 1984, 1999; Pollitt and Bouckaert 2004; Weaver and Rockman 1993; Wollmann 1997) exert pervasive influence over the “ways things are done” in particular
<table>
<thead>
<tr>
<th>Organizations Where Interviews and Documentary Study Were Carried Out</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Forestry</strong></td>
</tr>
<tr>
<td>Finland</td>
</tr>
<tr>
<td>The Netherlands</td>
</tr>
<tr>
<td>Sweden</td>
</tr>
<tr>
<td>United Kingdom</td>
</tr>
</tbody>
</table>
countries. Generic models and techniques (such as performance management) can therefore be expected to be heavily adapted—sometimes transformed—within different jurisdictions. This is a literature pitched mainly at the macrolevel of whole governments and political systems and one that stresses continuities. The way major institutions were set up and infused with particular cultural norms casts a long shadow down the years. Foundational decisions and the norms that grow up around them result in “path dependencies” in which the penalties for disturbing existing institutional patterns often exceed the rewards for doing so, even when short-term instrumental rationality indicates that change would be advantageous (Pierson 2000; Pierson and Weaver 1993; Pollitt and Bouckaert 2004). We term this body of scholarship “historical institutionalist” because it stresses both that “institutions matter” and that institutions, once established, often continue to exert an influence on decision making long after the original circumstances that gave them their particular shape have disappeared. However, we have to be careful in translating such broad influences down to the level of specific management practices such as performance measurement. For example, many of the findings of a study like Weaver and Rockman’s Do Institutions Matter? (1993) concern high-level policy making rather than the month-by-month management on which our research was focused. Thus we selected for special attention those elements that seemed likely to exert their influence not merely in the stratosphere of legislatures, cabinets, and party battles but also in the detailed steering of particular executive agencies. Bearing these considerations in mind we chose the following ideas as having strong potential relevance for our research.

First, majoritarian, single-party forms of government enable rapid change but also rapid reversals and alterations, both because there are fewer veto points than in, say, consensualist systems and because majoritarian party systems have a greater capacity to impose losses on particular interest groups (Pollitt and Bouckaert 2004; Weaver and Rockman 1993). Among our countries the United Kingdom stands out as having the most aggressively majoritarian system, with all three Continental countries having (in different degrees) multiparty systems, and the Dutch in particular being famed for their strongly consensualist orientation (Kickert 2000):

\[ H_1 \] Thus, we might expect public management reforms to be more radical, but also more reversible, in the United Kingdom than in the other three countries.

(On reversibility, however, American readers should note that, unlike their own “spoils” system, in all these four European states, agency bosses are career executives and usually do not change with a change of government.)

Second, countries with cultures that score high on uncertainty avoidance and high on collectivism may find the more risk-accepting and individualistic aspects of New Public management (NPM)–type reforms less acceptable than countries with lower uncertainty avoidance and higher individualism scores (Hofstede 2001; Mouritzen and Svara 2002; Pollitt and Bouckaert 2004; Pollitt et al. 2004). According to Hofstede (2001, 500), the Swedes are the most accepting of uncertainty, and the Finns, the least (Finland = 59, the Netherlands = 53, Sweden = 29, and the United Kingdom = 35). There are also considerable differences in individualism/collectivism scores, with Finland at just 63, the Netherlands at 80, Sweden at 71, and the United Kingdom at 89 (where the higher the score, the higher the individualistic tendencies):
H2 We might therefore suppose that the United Kingdom would be most comfortable with radical NPM reform, since it scores high on individualism and is relatively accepting of uncertainty. By contrast, the highly collective and risk-averse Finns might be more resistant to such reforms, with the Dutch and the Swedish somewhere in between.

The second theoretical tradition focuses on a different, more local (meso-micro) level of governance. It asserts that management is extensively context dependent (Whitley 1988, 1989) and that one key feature of that context is usually the nature of the primary task(s) in which a particular organization is engaged. The task gives rise to a number of work-related variables that have been widely recognized as important to organizational structuring for more than a quarter of a century (see, classically, Mintzberg 1979, 220–23). Thus, for example, a “production” task such as the issuing of driving licenses is standardizable, predictable, and measurable in ways that a “coping” task such as mental health counseling is not (e.g., Kaufman 1967; Lynn, Heinrich, and Hill 2001; Pollitt 2003; Wilson 1989). Another key contextual variable is the political salience of certain tasks—some attract endless political interference, while others languish in obscurity (Gains 2004; Hogwood, Judge, and McVicar 2000). A third important task characteristic is simply budget weight—some tasks are so expensive that they inevitably attract the attention of ministries of finance, whose task it is to macromanage the economy (Gains 1999). Others require only a modest budget and may lie “below the budgetary radar,” at least when public expenditure generally is not under great pressure. These task variables imply that the optimal approach to managing one task is unlikely to be optimal for another. We term this broad approach the task-specific literature and single out the following key ideas for testing:

H3 Ministries will be able to play a more active role in both setting targets and monitoring performance where the task is relatively simple, understandable, and measurable, for example, in Wilson’s “production agencies” (1989).

H4 Performance measurement will be either less important or more volatile for tasks that are highly politically salient (Hogwood, Judge, and McVicar 2000).

H5 “Big budget” agencies will be objects of elevated interest from both parent ministries and finance ministries because of their public expenditure consequences (Gains 1999).

If we consider our four functions against these three task variables, we could locate them as indicated in table 2.

**RESEARCH METHODS**

There was extensive study of the internal documents and systems of the agencies concerned. Eighty-four semistructured interviews were conducted with more than ninety senior managers, using a common interview schedule across the different participating organizations. Answers were inductively coded to provide the basis for the present analysis. Published (public domain) materials relating to agency performance were also analyzed, and their use was traced through other, related organizations such as “parent” departments, legislatures, audit offices, and so on. There was also a good deal of e-mail
communication and informal conversation with officials in the various agencies and ministries concerned, clarifying and updating points identified during the interview program. More details are given in the appendix. The agencies and ministries included in the study are shown in table 1.

**Table 2**
**Task Characteristics of Selected Agencies**

<table>
<thead>
<tr>
<th></th>
<th>Measurability/Standardizability</th>
<th>Political Sensitivity</th>
<th>Budget Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forestry</td>
<td>physical outputs measurable;</td>
<td>high in Nordic countries (forestry is big business);</td>
<td>big in Nordic countries; small to medium elsewhere</td>
</tr>
<tr>
<td></td>
<td>environmental impacts hard to measure; medium standardizability</td>
<td>medium elsewhere</td>
<td></td>
</tr>
<tr>
<td>Meteorology</td>
<td>forecast accuracy hard to measure; high science</td>
<td>usually low</td>
<td>small</td>
</tr>
<tr>
<td>Prisons</td>
<td>outputs easy to measure; impacts very hard to measure (“procedural” organization)</td>
<td>medium to high</td>
<td>medium to big</td>
</tr>
<tr>
<td>Social Security</td>
<td>“production” organization—easy to measure</td>
<td>high</td>
<td>very big</td>
</tr>
</tbody>
</table>

**The Nature and Use of Performance Indicators: National Comparisons**

It is important to state at the outset that both the idea and the existence of performance indicators (PIs) are anything but novel. All the agencies had operated with PIs for years, and in most cases these systems appeared to be undergoing continuous improvements. They had everywhere become part of the normal organizational discourse. The following response from a Swedish prison manager concerning local prisons is typical: “I think the discussion about why performance has been different than requested or expected has become much better. The reasoning of local organizations as to why their performance has not been good is improving.” Similarly, a Swedish social security and employment service manager said: “It’s more formal now, and it’s more that the organization throughout has the same targets.” Dutch managers at the Meteorological Institute (KNMI) and the prisons agency also supported the idea that “we are asking more managerial questions of our indicators.”

Indeed, most respondents acknowledged that there had been changes over time and that these constituted a shift toward more performance measurement and, to a lesser extent, more use of these measures by a wider range of managers: “There is definitely a change in atmosphere over the years; the results are more in focus than before” (manager, Swedish forestry agency). Similarly, a U.K. Benefits Agency manager claimed, “We are now moving towards a more holistic view of performance.” This was not only that attention had spread to nonfinancial as well as financial indicators. It was also that most of the organizations we visited acknowledged an increasing need to use client- or customer-based measures of quality. For example, “Customers are interested in total quality, not just the professional quality” (manager, Swedish meteorological agency).

A further finding pointing toward a degree of international convergence is that, when we asked, “Who makes the most use of performance data within the agency?” senior management usually cited themselves—everywhere. Sometimes parent ministries took
some interest, but often it was not much. Occasionally customers or pressure groups followed the data (e.g., environmental groups watched indicators of the management of state forests), but these were the exceptions rather than the rule. The only regularly active external (to government) scrutiny seemed to occur where state agencies were engaged in commercial activities in which they had private sector rivals or dependents (as, to differing extents, was the case in meteorology and forestry).

Against this generic background we nevertheless found some significant differences among countries. There were different attitudes to the process of performance measurement, with Sweden at one end of the spectrum and the United Kingdom at the other. Quite a few of the Swedish responses to our question, “What are the most important measures or indicators?” were extremely vague. Here is one manager in the meteorological agency: “I suppose it is what the government wants us to concentrate on, but this has not always been clear.” Swedish managers tended to describe the organizations’ performance in terms of maintaining interstakeholder relationships, whereas their British counterparts more readily focused on their specific targets and indicators. A good example is a senior manager at the Swedish forestry board, responding to the question about which PIs staff saw as most important: “With our field personnel it is important that they have equal relations with forest owners, that they are not seen as an authority.” Not only is the emphasis here on good, cozy relations, with no mention of any actual PIs, but there is also a denial of the fundamental role of the board, that is, as an authority that regulates the forest. British managers were much more precise about what their organizations’ indicators and targets actually were and also seemed to regard them as more consequential for their own standing as managers: “It [PI evidence] doesn’t drive people at the bottom of the agency particularly, but it definitely does the chief executive” (manager, U.K. Forest Enterprise); “the chief executive and his board use them [PIs] to modify performance proactively and to continually monitor and modify what they are doing” (senior manager, U.K. meteorological office).

The Dutch and the Finns mainly came somewhere between these Swedish and British poles but for the most part appeared closer to the Swedish pole than the British one. The tempo and tone of performance management were more relaxed. The introduction of a balanced scorecard system in the Dutch Social Insurance Bank (SVB) was typical. Two years of internal discussion followed by at least three years of carefully negotiated implementation constituted the timetable. Three years after the idea first came to the fore a manager was still telling us: “We still have a way to go spreading this performance information model throughout the entire organization.” Finnish agencies also seemed to move consultatively and cautiously. High-speed change has been much more common in the United Kingdom.

**The Nature and Use of Performance Indicators: Task Comparisons**

Meteorology is highly scientific, and forestry management is increasingly so. Partly in consequence, meteorological and forestry agencies are generally able to define their own nonfinancial PIs: “The ministry rely on us if we say this is a fair way to judge our performance, then they accept that” (manager, Swedish meteorological agency). Similarly, the Dutch meteorological office said: “We report to the assistant secretary general [in the ministry], but we suggest the measures she should use. . . . She is not in a position to argue with us.” A British manager from Forest Enterprise replied: “If you are asking who has the final say, then its probably ministers, but the measures come from this level [i.e., senior
management] and then carry on up the system.” Managers at the Finnish meteorological agency said that they suggested which measures should be used and that there was seldom more than minor adjustments.

Prisons are more of a mixed case. Politicians and the media are more sensitive to episodes such as escapes or riots and may insist that this or that aspect is given priority. Yet here, too, specialist knowledge has its influence: “I would say that we here have a lot of influence on what the government decides because we are the experts” (manager, Swedish prisons). In the Netherlands we were told: “Formally, the minister of justice gives us money and a task. . . . In practice headquarters [of the prisons agency] determine the performance measures in negotiation with the prison governors.” The reason, according to another manager, was that “the ministry doesn’t know what indicators are important or how they can be developed.”

The task of social security organizations is more straightforward—to pay legally defined benefits. These are the agencies where intervention by the ministries in target setting is strongest. Even in Sweden we heard: “We are not always satisfied about the extent to which we are involved in this process or the targets that are chosen by the government because, for example, this year a few of them contradict each other to some degree and they don’t give a clear picture of what should be accomplished” (senior manager, Employment and Social Security Agency).

KELA, the Finnish Social Security Agency, is a partial exception to this generalization, because, uniquely, it is under parliamentary not ministerial supervision. However, the KELA supervisory board has been strengthened. In the Netherlands, toward the end of our research, the body that supervised SVB was drawn back into the ministry, signaling a wish by ministers to exert tighter scrutiny. In the United Kingdom, too, parts of the Benefits Agency were taken back into the parent department, and our interviews indicated considerable interest by the ministry in setting targets. There had also been a recent performance audit by the National Audit Office (1998), putting a detailed critique of the indicator set into the public domain, which led to subsequent changes.

Part of the difference may be due to the fact that social security is not rocket science. Whereas ministers and generalist civil servants may well feel that they do not understand the science of weather forecasting or even the complexities of environmentally sensitive forest management, they are probably more confident that they grasp the essentials of handing out financial entitlements to citizens. As one KELA director said to us, their business is essentially simple: “Pay the right amount, pay quickly, be polite.”

Financial indicators usually have a higher status—in all four of the sectors we examined. Financial limits are definitely set by government. And whereas failing to meet nonfinancial targets may not be treated as a great crime, failures to hit financial targets are more serious. As one Swedish ministry civil servant put it, the discussion between ministry and agency “is generally focused upon money and not so much on performance.” Asked what the most important indicators were, a top manager in U.K. Forest Enterprise simply replied, “Unfortunately, the financial,” before going on to talk about social and environmental targets.

To our question about whether PIs really reflect the most important aspects of the agency’s performance (see appendix, question 13) we received a variety of answers, which varied somewhat by task. Prisons were perhaps the most problematic, and the managers were clearly conscious of the wider debate about whether prisons actually do any good in the longer term. Here is a not untypical response, from a Swedish prison manager: “I think
our annual report gives a good illustration of what we do and what is important. But as one journalist said after reading it: ‘The primary question has not been answered—that is preventing criminals from doing it again.’” A Finnish prison manager concisely expressed the interdependence of prisons and what goes on beyond their control, in society: “We have to address the social and self controls in order to have effective (and less harsh) officer control.”

In both the Finnish and the U.K. forestry agencies it was pointed out to us that, while measuring timber volumes and prices might be fairly straightforward, the measurement of environmental and sustainability concerns is both more recent and methodologically more challenging. One U.K. Forest Enterprise manager explained: “The environmental and social performance is difficult to get a handle on. We have no way of measuring these nonmarket benefits.” Nevertheless, measure them they do. Number of forest visits, number and duration of forest holidays, biological diversity—all these and more are measured, and the forms of measurement are getting steadily more sophisticated. But in no case could we find that these measures were directly used to arrive at the (inevitable) trade-off between commercial and environmental concerns. Here is the Forest Enterprise again: “There is nothing which says, ‘Do you sacrifice this to achieve that?’ . . . There is no formal equation.”

A parallel tension exists with research. Asked how he conceived “performance” for the U.K. Forest Research Agency, a senior manager replied: “Split performance. There is the fiscal side, and there is good quality research to customers. The two don’t go hand in hand sometimes.”

The Degree of Integration of Performance Information with Other Major Management Systems: National Comparisons

Here again there were differences among countries. Overall, the answers of Swedish managers were the most vague. We gained the impression that linkages between, on the one hand, nonfinancial performance indicators and, on the other, planning systems and human resource management (HRM) systems were often quite weak. In both the Nordic countries there is considerable resistance to performance pay, which is seen as divisive and potentially unfair. KELA said it would shortly introduce performance-related pay, but “we are meeting resistance from unions and the staff do not want it.” In Sweden the link between performance and finance was seldom strong: “In fact we see that there is no real connection between budget and the things that you must do, the targets to achieve certain objectives” (manager, Swedish prisons agency). Once more the biggest contrast was with the United Kingdom. Managers in Forest Enterprise explained that there were strong links between the PIs and financial planning and that senior managers were paid performance bonuses that were also tied in to the same system. On the planning side the links were significant, but overall integration was less.

The Degree of Performance Integration with Other Major Management Systems: Task Comparisons

A mixed picture emerged with respect to integration with financial management. Many respondents answered this question (appendix, question 15) by elaborating on how resources were allocated within the organization but without making any further reference to PIs as such. The Dutch KNMI straightforwardly stated that “performance indicators are
not linked to our financial data.” However, this was not typical. In most cases there was some link, not least because the primary performance targets were directly derived from the budget. All our U.K. agencies argued that the links between PIs and finance were close. The Public Service Agreements system introduced by the Labour government from 1997 had reinforced this and in doing so had given the Treasury an even stronger grip on departments’ and agencies’ delivery of targets.

A senior Swedish prison manager thought that PIs and planning (question 16) were closely linked: “Our planning system is very well integrated with the use of performance measurement, but only where measurement is possible: issues such as recidivism [are] less amenable to our planning processes.” Most U.K. agencies claimed that the planning/performance targets linkage was close.

Generally links were weakest with HRM systems, especially in the two Nordic countries (question 17). Other factors seemed to take predominance in pay and grading decisions. In the Netherlands the KNMI said, “We view personnel issues quite separately from our performance indicators.”

Ministry/Agency Relationships: National Comparisons

Both the Finnish and the Swedish respondents tended spontaneously to introduce the notion of trust and to stress what one might term the “coziness” of the relationship with the ministry. This kind of sentiment was rare among the U.K. interviews and not that common in the Netherlands. In no single case in Sweden or Finland did we receive the impression that the ministry was pushing very hard on measured performance (see, e.g., Joustie 2001). One official at the Finnish Ministry of Finance (which has a department responsible for public management) put it like this: “They [the parent ministries] can play this game quite nicely. The normal way of functioning is that they receive the report [from the agency], and they put it on the shelf.” Another Finnish official explained, “Our mantra is that ministers should only be concerned with strategic steering,” but this hymn, while often sung by central departments, found little echo in the agencies we investigated. If there was any strategic steering, it took place more through new legislation, budget setting, and reorganizations (i.e., traditional ministerial instruments—and not the focus of this research) than through the diligent use of performance data. Indeed, the tough part of the dialogue with ministries was seen as forming the budget, and very few respondents linked that to PIs—it was much more about current macroeconomic and fiscal policies combined with the political issues that were the current foci of ministerial and media interest.

So “steering by indicators” was often light. For example, in Sweden the annual regleringsbrevet—the policy letter from the ministry to the agency—tended to be cast in general terms, with the formulation of actual measures and indicators a matter for subsequent negotiation. Typical, perhaps, is this description from a senior Swedish prison manager: “We have frequent discussions with the ministry during the process of setting targets, we have very good relations with them. This is very important because they need to know what is a reasonable request to ask of us.” The perspective from the ministry side of this same relationship is quite interesting. A civil servant told us: “Perhaps we ask the wrong questions, because although we get lots of answers we do not get the answers that we need—the answers that tell us about the results.”

The culture of the Nordic countries also militates against harsh measures if targets are missed. When we asked an official in the Finnish parent ministry what would happen if the
Finnish Meteorological Institute fell significantly short of a target, he said: “It might lead to a change in the wording of the target. But actually it seldom seems to happen... But nowadays the ministry is more willing to set higher targets than it was.” Similarly, when asked what would happen if the results for a particular prison were very bad, a Ministry of Justice official said: “We don’t have very good ways to react. In Finland people are not sacked over that kind of issue.” Even in the more commercial environment of forestry enterprise, the ministry acknowledged: “Normally we have no hard ways to use against them if they don’t achieve their targets.”

In the case of the Netherlands, ministry/agency relationships were not described in such intimate terms, but neither was there much sense of active steering by ministries. Here is a KNMI manager:

Well, it is formally... we are steered on output... but in my period I never had to give responsibility for my activities. At the end of the year we had to give a report about the outcome of the institute, and I never had any debate about that... There was not really a culture, as there was in England, on the results... Then we also had no discussion about priorities. You had to do it yourself. So that was not really steering, that is what I am saying.

This same (very experienced) manager concluded with a particularly striking observation: “I must say that in my view the ministry is not much, by its nature, interested in performance. The ministry by its nature is interested in problems. They have a complete other set of priorities.”

As noted above, specifically financial targets are frequently mentioned as both tough and “imposed” from above. Indeed, for Dutch agencies, their mode of financing seems to have become the dominant factor in their very conceptualization as a new type of public body (Smullen 2004). In many cases the chief financial inquisitor is seen to be the Ministry of Finance rather than the parent ministry. Here is a civil servant from the Swedish Justice Ministry, which is concerned, inter alia, with the police and the prisons: “We propose the budget for the agencies to the Ministry of Finance. But we also have very tough negotiations with them... I think that the Ministry of Finance has quite an influential role in this discussion.”

A manager with U.K. Forest Enterprise, caught in a balancing act between hard commercial and soft environmental and social targets, put it more brutally: “The politicians are not the problem, it is the Treasury.” Similar sentiments came from some of the managers at the Finnish Metsähallitus, who saw an ever larger slice of their profits being clawed back by the Ministry of Finance. A U.K. Treasury official we spoke to echoed a characteristic central department view that we also heard expressed in Finland and the Netherlands: “Many departments don’t take target setting seriously enough. Often targets are too loose.” Similar words were heard from the Cabinet Office.

The Effects of Agency Status

In Finland and Sweden, agencies were a prominent and long-standing component of the system of government. For the Netherlands and the United Kingdom they were relatively novel. In these countries, therefore, we sometimes asked a supplementary question about the effects of recently acquired agency status: “I would not say that our contact with the ministry has increased or decreased since we became an agency. KNMI has never experienced a great
deal of political pressure: we do popular work” (manager, KNMI). Another manager from
the same organization put it most bluntly: “I think this is quite an autonomous organization
and that is not influenced by being an agency or not being an agency.”

It would be incorrect, however, to suggest that there are no effects at all from
agencification. A Dutch prisons manager explained: “I think perhaps there was help from
the change because it has made our core business in both the central office and in the field
clearer. We [the head office] are responsible for making clear the goals of the organization,
and they are responsible to execute the prison sentence.”

In the United Kingdom there were fairly sober assessments of the advantages and
disadvantages of agency status. While financial and personnel freedoms were real and
appreciated, the policy/operations split was the focus for some criticism, both from parent
ministries and from agency management. Here is a senior member of the Forestry
Commission—the department that supervised the two U.K. forest agencies: “There was
too much emphasis, initially, I think, on people appointed to run agencies thinking they
could run off and do their own thing and leave the department behind . . . The Whitehall
mandarins have become increasingly mandarin-like, even more divorced from where the
policy is being applied” (see also Office of Public Services Reform 2002).

In one case—the Finnish prisons and probation agency—we had the opportunity to
interview staff just before and shortly after transition from being part of the Ministry of
Justice to being an agency. The transition certainly was not dramatic. The ministry said it
wanted to move toward results-oriented steering but acknowledged that nothing much had
changed yet.

**Ministry/Agency Relationships: Task Comparisons**

In all four countries three factors were frequently mentioned as influencing the behavior of
parent departments. The first was the intelligibility of the task to generalist politicians and
generalist or legally trained civil servants in the ministries. Ministries are understandably
more cautious about intervening in the details of activities that they realize (or believe)
they do not fully understand. The observability and measurability of outputs and outcomes
are vital supports for understandability by nonexpert “outsiders” (Wilson 1989). In
addition, one factor that was quite frequently mentioned by our respondents was the sheer
“technicality” of the task. Weather prediction, for example, produces lots of measures and
is, in a sense, eminently observable. The interpretation of those measures, however,
requires considerable scientific knowledge. On the whole ministries think they understand
social security, and they know what prisons are (although the lack of outcome measures
sustains an ongoing debate about their ultimate usefulness). Forestry is, in principle,
straightforward, but increasingly there are craft and scientific aspects (e.g., biodiversity)
about which the ministry will find it difficult to argue with the agency experts. Meteorology
is high science—involving folk in white coats with big computers and an excluding
technical vocabulary. How one measures the accuracy of a weather forecast may seem
simple to a layperson, but a five-minute conversation with a meteorologist will convince
one otherwise.

A second factor was the political salience of the primary task. Meteorology is
normally noncontroversial, barring climatic disasters. Forests are usually fairly quiet in
terms of party politics, although in the Nordic countries they are big business and therefore
the targets for strong commercial lobbying. Prisons are much more in the daily eye of
politicians and the media. Prominent escapes are news, especially if the escapees go on to commit further crimes. Whether the prisons are becoming too “soft” has been an issue in all four countries, as has the abuse of drugs by inmates. Even in Finland, with low crime rates and a tiny prison population by U.K. or U.S. standards, we were told that “the atmosphere has changed. Politicians want more punishment. They don’t accept our liberal ideas.” Thus prisons tend to be subject to spasmodic but strong interventions from parent departments. Social security is also high on political agendas, partly because it is big budget and partly because it directly affects so many citizens’ daily lives but also because of political and media campaigns about “scrounging” and fraud—especially but by no means exclusively in the United Kingdom.

The third factor was simply size of budget. In meteorological offices we were frequently told that their ministries did not bother them too much because, in budgetary terms, they were “peanuts.” By contrast social security is a big budget item, and ministries of finance as well as parent departments are often looking for economy (although in the unusual case of Finland, KELA is subject to parliamentary rather than ministerial direction).

The following examples illustrate these three influences: “If I may say so, they have no expertise in this field, and they trust us because they can’t really investigate it themselves” (manager, Swedish meteorological agency); “at the ministry hardly any attention is paid to these indicators. Even if performance was bad I suspect they would not react unless we pointed their attention to this” (Dutch meteorology office); “they [the ministry] ask . . . why we have not achieved a certain level in an indicator, but they are not in a position to query or argue our answer because they do not have alternative benchmarks for the answer we have given them” (manager, Dutch prison). Where an agency is small budget, high science, and not politically controversial, it can even come to feel neglected. This was the response from U.K. Forest Research: “I would have to say that the amount of use [of performance data] made outside the agency is really quite modest. It is one of the things I find a little bit distressing, inasmuch as forestry commissioners and ministers have not commented on our targets.”

An additional factor is that there was a very clear difference between agencies that had a substantial commercial component and those that did not. Here is a senior member of the U.K. Forestry Commission: “There is a lot of mystery spoken about agencies, but they are just a way of applying commercial principles to running a function. . . . It is about giving people a focus, a balance sheet, good indicators of progress which are measurable and judging them against those.” Even Finnish meteorological office managers, when asked what the most important external influence on their performance was, replied: “Since 1995 the world of competition has come to us.” A strong commercial dimension thus gives agencies added autonomy from ministries in two ways. First, it makes them less reliant on the parent ministry for revenue (indeed, in cases such as the Finnish Forestry Enterprise, the ministry is reliant on it for revenue). Second, it gives agencies an argument that is hard for the ministry to counter: “It was the market.” At both the U.K. and the Finnish forestry organizations managers acknowledged that financial targets were sometimes missed because of market factors beyond their—or their ministers’—control. Asked what the most important external influence on his organization’s performance was (a question that with noncommercial agencies usually brought an answer of the parent ministry), a Forest Enterprise manager had no doubt: “The world timber market. We have to ride the waves and cannot influence it.”
DISCUSSION

When we look at the evidence about performance management and performance steering across four countries and four sectors, patterns can be discerned, but they are complex. To a considerable extent the general expectations of similarity and difference we derived from the historical institutionalist and task-specific literatures were confirmed, but there are also some overarching elements of convergence that neither body of theory fully accounts for.

The observed international differences fit well the first two ideas derived from the historical institutionalist theory (see theoretical background above). U.K. administrations can push performance measurement faster and further because they operate within a majoritarian, one-party system of government, with fewer veto points and more possibility for imposing “losses” on powerful groups (in this case, especially career public servants). What is more, the more individualistic and risk-accepting culture allows the more vigorous use of devices like performance-related pay and transparent public reporting of targets and achievements. The other three countries cluster together as more consensualist and (in the cases of the two Nordic countries but less so in the Netherlands) more collectivist. Performance measurement is accepted as a legitimate modern technique, but it is used within a more negotiative, consensus-seeking cultural climate. The three Continental states also have quite strong corporatist features, which means that policies are usually developed informally, through networks of key institutions that bargain within a consensualist culture (Kickert 2000; Pierre 2004). In such circumstances hard-edged performance steering may threaten important relationships and interests—not least those of the powerful public service unions or political parties that are allied to them or to certain agency missions. This may well be part of the explanation for the very slow application of performance-related pay in the Nordic states and the seeming lack of integration between performance management and human resource management. Furthermore, the particular “vagueness” noted among Swedish respondents may be explained by the extraordinary strength and longevity of the agency system in that country. The interpretation would be that Swedish ministries—to a greater extent than in Finland or the Netherlands—frequently lacked the capacity and status to challenge agency autonomy (see, e.g., Molander, Nilsson, and Schick 2002; OECD 1998).

The United Kingdom, by contrast, is a highly centralized, majoritarian system in which ideas and practices, once they have been adopted by the central executive, can be spread quickly and forcefully throughout the public sector (Pollitt and Bouckaert 2004). All these long-term institutional factors have made it easier for U.K. governments to push for earlier, more rapid, and more widespread implementation of performance management than their Continental counterparts. While reforms have indeed been subject to shifts as governments have changed, the basic direction has not altered because both Conservative and New Labour administrations have embraced the basic NPM doctrines, even if they have wanted to change the labels on particular reforms. Also, in the United Kingdom PI systems are more intimately connected to the other main management systems (financial management, planning, HRM) than in any of the Continental states. Finally, it is here too that we find examples of the most energetic attempts by ministries to use performance indicators as a basis for steering agencies—although even in the United Kingdom such attempts could best be described as spasmodic. This “U.K. exceptionalism” seems unlikely to be due to the United Kingdom starting earlier and being further along the learning curve. Performance measurement has been a major policy priority in all three
Continental countries since at least the late 1980s (Pollitt and Bouckaert 2004), so they have all had plenty of time to catch up.

The patterns of evidence are complex not only by virtue of intercountry differences but also because they seem to vary significantly by primary task. As we have seen, measurability/standardizability, political salience, and budget weight all seem to play their part in shaping the direction and form of performance management. Each of these task-related factors has been identified in previous work on agency differences, and to that extent this research confirms earlier theorizing (Gains 1999; Wilson 1989). Further, they could perhaps be connected with the “clarity/discretion matrix” recently proposed by U.S. performance management researchers (Ingraham, Joyce, and Donahue 2003, 126). They combine in various ways to produce contexts with high or low clarity concerning missions and high or low levels of discretion for managers.

There is space here only to offer the briefest indication of the action of these task characteristics. The effects of measurability and standardizability are clearly seen in the case of social security (high on both). Ministries have intervened more and more often in the performance management of social security agencies than in any of our other three tasks. Political salience—our second key characteristic—is perhaps the most powerful and immediate influence and the most unpredictable in its effects. The intense political orientation to “problems” seems virtually universal. In the prisons sector it is vividly encapsulated in the saga of Derek Lewis (1997), a businessman brought in by the Conservative administration to head the new U.K. prisons agency. Following the much-publicized escape of some high-profile prisoners, Lewis was sacked by his minister. Lewis sued for wrongful dismissal, pointing out, inter alia, that he had achieved all his performance targets, including those related to rates of escape. The matter was settled out of court, to Lewis’s considerable financial advantage. For our purposes the moral of the story is that, when embarrassments, scandals, or disasters occur, politicians and the media suddenly take an enormously detailed interest in organizational activities they have never asked about before. This interest includes performance data, but those data are unlikely to be given any privileged standing during bouts of political conflict—more probably they will just be treated as extra ammunition.

Our third task-related expectation was that large-budget agencies would attract more attention from ministries than small-budget agencies. By and large this seems to be borne out. Social security and prisons are the two largest net budget functions, and they do indeed seem to attract more attention from ministries (especially finance ministries) than forestry (usually commercially profitable) or meteorology (small budget). However, this finding is subject to the point that these are also the two most politically salient functions in our set, and we are unable to disentangle the relative contributions of the two influences.

Thus far our theoretical expectations have been largely confirmed. What our chosen theories do not account so well for are convergences across both countries and tasks. “Performance” is becoming a more widely used and widely understood concept in all the contexts that were examined during the EUROPAIR project. It is everywhere an expanding concept: from previously having had a predominantly financial character, it has acquired greater multidimensionality, incorporating not only “producer” concerns such as efficiency and cost-effectiveness but also “user” elements such as service quality and customer satisfaction. All these types of measure have become “normal” in the sense that the case for them is widely accepted. Agencies in all countries were using or beginning to develop versions of the “balanced scorecard,” signifying a belief that performance henceforth needs
to be conceived and measured multidimensionally. This kind of international, intertask convergence seems capable of sweeping over national and sectoral differences—it is not necessarily a very deep wave, but it is certainly wide in its reach. To this extent the theorists who posit broad international trends and tendencies—lubricated by the growing ease of rapid international communication and the work of international organizations such as the OECD—are confirmed (e.g., Kettl 2000; OECD 1997; Pollitt 2002).

A second generalization is that nonfinancial measures still tend to take second place and that expanding the “vocabulary” of performance measurement, along lines suggested nearly twenty years ago, is quite a slow-moving business (Pollitt 1986). Balanced scorecards are becoming common, but in ministry/agency relationships the bottom line is still usually getting and sticking within the budget, and the links between performance and money inputs frequently remain weak or unclear. These European findings align with recent large-scale research in the United States, where the researchers concluded:

Many governments have gotten the performance measurement message, in the sense that they have moved aggressively towards identifying outcomes for their programs and measuring progress towards them. The next great challenge for them is using that information to make decisions and policy. In particular, the use of performance information in the budget process is the next significant step in the movement towards performance management. (Ingraham, Joyce, and Donahue 2003, 155)

Despite these partial disconnects, the technical instrumentarium of performance measurement was under constant, incremental improvement in virtually every organization we visited. The findings of the U.S. Government Performance Project that, first, “integration is neither an obvious choice for governments nor an easy process” and, second, “an integrated performance-based government is a progressive, building-block process” were both true for the European experience too (Ingraham, Joyce, and Donahue 2003, 132–33). Indicators were being refined, and the validity and reliability of data were subject to steadily more and more sophisticated checks. However, most checks were internal: few agencies subjected their nonfinancial performance data to regular external validation. Financial data, however, are routinely subject to audit.

While this extension and strengthening of performance measurement may be the main direction of international “flow,” it is, as indicated earlier, a tide that runs more or less strongly in different parts of the ocean and breaks upon organizational coastlines with rather variable effects. It is clearly in the United Kingdom that performance management has had the longest history and the most vigorous expression. The three “Continental” countries are all more cautious, more consensual, and more consultative in their attempts at performance management.

These complex patterns of simultaneous convergence and divergence point toward several opportunities for further research. International convergence theories have already received a good deal of academic attention, and our research is not the first confirmation of the importance of national trajectories and styles. Nevertheless, there is an obvious thinness in the work on cultures. Almost everyone agrees culture is important, but there is little systematic empirical research that focuses specifically on public sector organizations and even less that is comparative. We have had to rely too heavily on the now rather old and very general work of Hofstede, and this imparts a fragility to this aspect of our interpretations.
Task-specific factors have been less to the fore in recent public administration literature, and attempts to combine a mesolevel analysis of tasks with a macrolevel analysis of national systems have been few and far between (although see Batley and Larbey 2004). Our comparative analysis of countries and tasks indicates that the applicability of particular models—such as performance management—is heavily contingent on the interaction of institutional patterns, institutional norms, and specific task characteristics. These create contexts in which management has to work with clear or ambiguous missions and high or low discretion. As yet, however, the theoretical maps of these complex interactions are seriously incomplete. This study suggests that the “intelligibility” of a task to busy lay ministers and ministry officials may be a significant factor and that intelligibility cannot be entirely reduced to Wilson’s observability of outputs and outcomes (although these factors often covary). There is also room for more research on how executives in politically volatile tasks such as social security manage to manage—our efforts have only scratched the surface of this issue.

To offer a final encapsulation of the findings themselves, one might say that, in northwestern Europe, performance measurement has become almost universal. This is the level on which convergence is most evident, and it goes well beyond rhetoric and into practice. Performance management is growing steadily but varies in form and force among different countries and different tasks. Performance steering of agencies by ministries is still a rarity and may never become particularly common. Performance measurement and performance management remain activities conducted chiefly by and for managers. On the whole politicians do not take much interest, and neither do citizens—unless and until disasters, scandals, or breakdowns come along.

APPENDIX: THE EUROPAIR PROJECT

EUROPAIR was a research project focused on performance management in agencies and other “autonomized,” “arm’s-length” public bodies (Pollitt et al. 2004 provides a general account). A common interview schedule was used in almost all the interviews with agency personnel, but not all questions were used at every interview. Sometimes a selection would be made, depending upon the time available and the particular responsibility of the interviewee. Some interviewees chose a nonattributable basis, and others permitted us to identify them in our published quotations. Most were also offered the opportunity to correct the record.

This article is based on a particular subset of the questions asked (hence the gaps in the numbering below). Answers were retrospectively coded into a limited number of possibilities for each question; for example, the answers to question 10, which concerns who used the performance data outside the agency, were coded as 0 = no answer, 1 = parent ministry, 2 = finance ministry, 3 = customers/clients, 4 = no one much, 5 = other.

How Is “Performance” Conceived of and Measured in the Sample of Agencies Studied?

3. If I say to you “the performance of X [agency name here],” what are the first aspects that spring to mind?

4. What are the most important measures or indicators, from the point of view of agency staff?
5. Have there been any significant recent changes in the way in which performance is thought about here?

6. Have these changes been reflected in the measures and indicators?

7. Who determines what the current set of performance measures should be? (Where appropriate, pursue this with supplementary questions aimed at establishing how far the agency effectively sets its own indicators or, alternatively, how far indicators are imposed on it from outside.)

How Far, by Whom, and for What Are the Performance Data Used?

8. Are performance data much discussed within the agency?

9. Who makes the most use of them inside the agency?

10. Who (if anyone) makes the most use of them outside the agency? (Supplementary: Can you give an example of their use?)

To What Extent Are Performance Measurement Systems Integrated with Financial Management Systems? Where This Integration Is Low, What Are the Reasons for That?

13. How far do you think the performance data reflects the really important aspects of the agency’s ‘real’ performance?

14. In general, how closely is the performance measurement system linked in with the other main management systems in the agency?

15. What kinds of links are there with financial systems? (If weak, then why?)

16. What kinds of links are there with planning systems? (If weak, then why?)

17. What kinds of links are there with human resource management systems? (If weak, then why?)

External/Contextual Influences on Performance

24. What are the main external influences on how well the agency performs? (Supplementary: Through what process do these influences operate?)

REFERENCES


