

Performance management that works? Improving  
public services by making use of intrinsic  
motivation

Ian Greener

University of Strathclyde

# **Performance management that works? Improving public services by making use of intrinsic motivation**

## **Abstract**

One of the central features of public governance in the 2000s was an extension in the use of performance management (PM), but research has suggested that the ‘gaming’ of PM systems became commonplace and that measured performance data was often unreliable. However, PM is necessary in some form, and has fared better in some settings than others. This paper presents a systematic comparison based on two case studies where success has been claimed, from English local government and the Quality and Outcome Framework (QOF) in the NHS, as a means to generating insights into how PM systems can be made to work better. It links its analysis to Frey’s and Deci’s work on intrinsic motivation in order to theoretically explain its findings before exploring how the insights generated can be applied to other service settings.

## **Introduction**

Despite the significant criticisms made of performance management (PM) systems in public organizations, they are with us to stay. It is therefore important to find ways to try and improve PM. To that end, this paper examines critical cases where at least some success has been found by research, to consider how it might offer guidance as to how PM systems might be improved more generally. This paper explores two such cases from England, looking for similarities and differences between them, and locating the findings within the work of Frey and Deci to attempt to make them generalizable to other public service settings.

The paper first provides a context for exploring the problems PM systems have encountered, before moving on to explain its method. The two cases are then examined separately, before being compared them within an intrinsic/extrinsic motivation framework which was chosen as

the best candidate to theoretically deepen the paper. The paper concludes by explaining its relevance for performance management systems more generally. The paper therefore aims to make contributions in showing *how* PM systems can be improved, as well as trying to explain *why* systems designed along these lines stand a better chance of working.

### **PM systems in public organizations**

In the pre-New Public Management era, academics and managers were well aware that it was necessary to plan and measure performance (Chapman & Dunsire, 1971). However, there was a pragmatic recognition that the outputs of public services were often extremely difficult to measure, leading to doubt as to whether it was possible to manage public services according to clearly-defined targets (Carter et al, 1992). PM has always been with us in one form or another, but the use of targets and indicators as both measures and incentives, is a more recent phenomenon.

The New Public Management (NPM) represented a shift in the underlying philosophy of the management of public services that challenged the assumptions of traditional public administration (Dunleavy & Hood, 1994). Although when examined comparatively, the NPM has been introduced and implemented in very different ways from country to country (Pollitt & Bouckaert, 2011), its different models share a concern for improved efficiency and the measurement and control of the activities of public organizations (Bloomfield, 1991). In the 1980s, public services were urged to become more ‘business-like’, and in the 1990s the ‘reinventing government’ (Osborne & Gaebler, 1993) movement urged public services to become more customer-focused and concerned with outcomes. The growth in the capacity and usage of IT systems made it possible to measure and manage the activities of public organizations more quickly and directly than before, making PM a central component of the

NPM.

### ***The Trouble with PM***

At the same time as it has increased in usage, PM has suffered a range of problems.

First, there is the already-mentioned concern that public service outputs and outcomes are often very difficult to measure (Klein, 1982). This leads to the second problem that the measures chosen in PM may not capture the underlying phenomenon they are meant to (Lowe & Wilson, 2015). Ambulance services, for example, may come to be judged in relation to their expected maximum response times rather than whether they are actually getting to those who need their services quickly enough to be able to help them (Heath & Radcliffe, 2007). A third problem is that, in some contexts, PM systems have led to managers spending their time manipulating measures rather than trying to improve system performance. This ‘gaming’ seems to have been especially prevalent in hospitals in the National Health Service in England (Bevan & Hood, 2006).

Fourth, PM can be seen as creating a dilemma for managers between managing the system on a short-term basis to hit targets or looking to the longer-term which may lead to different priorities. This can lead to managers losing sight of the actual purpose of the PM system in the first place (Mannion et al, 2001). Fifth, PM systems are often put in place by policymakers at the same time as they impose top-down structural changes upon public services, resulting in uncertainty for partner organizations, disruption to services, senior staff turnover and reduced, rather than improved, performance (Andrews & Boyne, 2012). Sixth, PM is subject to a range of problems in setting targets levels; if too low a target is set, everyone will achieve it and it will have no credibility; if too high a level, policymakers will be concerned that their management of services will appear poor (van der Geer et al, 2009).

As a result of the emergence of the problem above, PM systems have lost a great deal of credibility. Governments have periodically promised to reduce the number of targets in the use of public services, even if they have struggled to honour those promises in practice. But despite the poor reputation that PM presently has, however, there are instances where some success has been claimed. These examples can be considered as ‘critical’ case studies (Flyvbjerg, 2001) in that they have the potential to teach us a great deal about the principles by which PM systems can be improved.

## **Method**

This paper utilises Pawson’s realist review (Pawson, 2006 , Pawson et al, 2005) which aims to understand how and why different systems work in different contexts to generate theories that explain their logics, and so is especially suited to considering cases of PM systems comparatively. To locate our critical case studies, we searched a range of on-line bibliographic and publisher databases to find empirical sites where there was sustained research on PM, and then examined that research to find work which claimed to have found success in its implementation. This process identified very few candidates, with the overwhelming majority of research conducted on PM systems instead identifying the range of problems with such systems outlined above. We then ‘snowballed’ from the two cases we had identified, following citation chains for work exploring the Comprehensive Performance Assessment (CPA) system in English local government, and the Quality and Outcomes Framework (QOF) in GP surgeries in the NHS, but did not find further instances of bodies of work which pointed to any kind of sustained success. A further search on the two cases showing some dissenting voices in terms of the success of the systems, but the work of Boyne and his collaborators in the case of CPA, and from a range of researchers in the case of the QOF in the 2000s, suggests there the potential for generating theory which might be able to explain the successful elements of the PM systems.

The research identified in each case was then examined in line with Pawson's approach, with each paper or report being summarised in line with its ability to contribute to an emergent theory of how each PM system worked (see also Tan & Harvey, 2015). These emergent theories were then systematically compared, and a theoretical account of the workings of the two systems constructed based on a comparison of the similarities and differences between them. This emergent theoretical account was then compared with different candidate theories in the literature which examine the link between motivation and public management, with one based around the work of Deci examining intrinsic and extrinsic motivation being chosen because of its ability to add additional depth to the findings from the cases, in line with realist research principles (Danermark et al, 2001).

We now turn to the two cases.

### **Performance Management in English Local Government**

The first case where success in PM has been claimed occurs in English local government. In England there is a unitary system of government in which the central government exercises a strong degree of both legal and financial control over local government, providing nearly three quarters of its funding through central grants, and limiting the scope of local government to raise funds independently. Local government provides a range of services including education, environment, and refuse collection.

There has been a PM regime in English local government since the 1980s. The range of performance measures has grown in number and complexity, reaching their peak in the Best Value regime introduced by New Labour in 1999 (Boyne, 2002). This regime was quickly acknowledged, even by central government, as being unmanageable, and in 2001 the CPA put in place to create a system of inspection and measurement that categorises performance into

five measures ranging from poor to excellent, but which was superseded by a star rating system in 2005/6. The system made as much use of pre-existing reports and data as possible, combining it into a range of scaled scores. Several dimensions of performance were assessed including organization activity, resource use and service provision, to generate a final score.

The introduction of the CPA came from the Audit Commission which both designed and implemented it, and which consulted with local authority leaders to achieve a consensus on its form so that there was little opposition to its introduction. Local authorities were given representation on the Audit Commission and included in the peer-reviewed inspections that formed the basis of the scores awarded, leading to them having some ownership of the process (Bertelli & John, 2010). The performance measures were published so that poor performers could be 'named and shamed', especially in local news reports, as well as through professional networks. Central government put in place a number of interventions they could use to deal with measured poor performance, including local councils being put into 'special measures' which entailed more active central government regulatory involvement.

The results of the CPA also seem to have democratic consequences. Poor performing local authorities often lost subsequent elections, indicating there was a real risk to those in elected positions for their assessed performance falling into one of the bottom two measures. Senior officers (including chief executives and service leads) especially regarded a poor result as a 'wake-up' call after which they had to show quick improvement (Boyne et al, 2012).

Performance measures in the CPA were broad-brush but clear, and local authorities had discretion in how they organized in order to meet their targets. Those measured by the system as being good performers received the promise of new funding streams, increased autonomy from inspection, as well as favourable local media coverage. The additional funding or

freedoms, however, didn't often appear. Despite this, CPA still appears to have motivated local government managers because of the competition it engendered between local authority areas. To quote a key source, 'CPA is part of the gossip-world of interested professionals, where both good and bad performance gradings affect the reputation of organizations and the careers of those working in them.....' (Boyne et al, 2012) (p.210). Bertelli and John (2010) demonstrate that, even though there was a political aspect to the rankings achieved by local councils, scores improved over the period of the CPA, with the average performance increasing year-on-year.

Despite its apparent success in driving the improvements outlined above, the CPA system was abolished in 2009 and replaced by the Comprehensive Area Assessment (CAA). It appeared that the CPA had been a victim of its own success in that so many councils had shown such significant improvement that it was expanded to include local partners of the council, as well as attempting to refocus more on outcomes and to be more forward-looking. However, doing so meant such significant changes that the move to the CAA involves the effective introduction of a new system.

The CPA was not without its problems. Haubrich et al (2008) build on earlier work to show that local authorities in poorer areas tended to get lower scores, even controlling for other factors, so penalising them for factors outside of their control. Andrews, Boyne et al (2005) also suggest that local authorities confronting very diverse populations often struggled to perform well, and point to the potential for some councils being 'falsely lauded' (p. 654) for having the good fortune to be 'operating in favourable circumstances' while others were 'wrongly criticised for the performance effects of difficult conditions' (p. 654). However, some councils in relatively deprived areas did manage to achieve considerable improvements, indicating it was possible to break the general trend.



McLean et al (2007) go further in suggesting that the CPA failed four key tests. They suggest a reliable and valid measure of performance must be ‘Invulnerable to categorization errors. Invulnerable to gaming. Consistent with other government policies. Uncorrelated with uncontrollable factors’ (p. 117). They suggest categorization errors occurred in the summarisation of data from existing sources into the CPA resulting in over-simplification and political decisions about the boundaries of the initial classifications. They find gaming effects such as output distortions where local authorities found ways of reducing services but increasing scores. They also found contradictory incentives in place between the extent of deprivation in place in a local authority area and its CPA score, and that CPA scores were ‘positively affected by discretionary expenditures incurred by local authorities’ who were able to ‘buy’ better CPA scores (p. 116). These are important concerns. However, whether it is realistic for any performance management to meet the very strong tests they suggest is questionable – no system is likely to be ‘invulnerable’ to categorization errors and gaming, or to be entirely ‘uncorrelated’ with uncontrollable factors. These standards are so strong as to be unreasonable, The key question is whether the CPA system dealt with these factors sufficiently to be considered successful, and it is striking that leading commentators such as Boyne (most clearly in Boyne et al, 2012), who McLean et al acknowledge as holding a very different view to them (p. 114), do appear to believe this to be the case.

There are therefore debates about whether the CPA was a success, but as there is clearly research which claims this to at least some degree in an academic literature which is overwhelmingly critical of PM, there is potential for positive learning. Outside of isolated examples, gaming seems to have been limited; senior local government managers appeared to regard the system as having legitimacy because of their involvement in designing and implementing it; despite rewards for high performance not appearing, penalties for low performance drove improvement because of competition between local authority managers to

secure higher ratings than their peers, and the fear of a bad result in a local election should their performance fall into one of the bottom two grades. In the general context of the difficulties PM systems faced in the 2000s, the CPA system offers us clues of how PM systems might be made to work better.

We now turn to our second case, that of the Quality and Outcomes Framework (QOF) in the National Health Service.

### **Performance Management and the QOF**

The QOF was introduced as a central part of the General Medical Services contract negotiated during the early 2000s and introduced in 2004. Within it, GPs were required to measure their practices against a range of performance indicators which were justified by policymakers through the links between the indicators and clinical evidence. The QOF awarded GP practices points based on their level of achievement across four domains (clinical, organisational, patient experience, and additional), with over 80 indicators in 19 clinical areas.

The introduction of the QOF, as with the CPA in local government, was the result of extensive consultation between central government agencies and professionals. The introduction of performance indicators was not new to general practice, but part of an incremental process of change going back to earlier changes in the 1990s which both introduced targets into primary care and emphasised an increase in practice nurse involvement in GP surgeries (Gunstone, 2007). GPs, as with local authority leaders, appeared to accept that they needed to adopt a greater focus on measurable outcomes (Exworthy et al, 2003).

The professionals most affected by the QOF seem to have accepted its introduction, with little opposition from either GPs or practice nurses (Checkland et al, 2007). Concerns that the QOF would increase workload have been justified, but this increase in workload does not appear to

have led to a reduction in care provision, with GPs taking on the extra work associated with it in addition to their clinical work (McDonald et al, 2009).

GP practices had considerable discretion in how they met the QOF's targets. Some practices used it as a means of triaging care so that GPs see the most clinically complex cases only (Sheaff, 2009) so that more routine work has been delegated to nurses and health care assistants (Leese, 2007), and with nurses reporting that they enjoyed the greater level of autonomy this offered (McDonald et al, 2008). There is also evidence that the changes led to an increase in teamworking between doctors, practice managers and nurses (Lovett & Curry, 2007). GPs, remarkably considering the history of antipathy of doctors to managers or managerialism, also used the QOF to challenge the performance of peers where they believe them to be endangering the practice's measured success (McDonald et al, 2009 , McDonald et al, 2008).

In terms of getting GP practices to meet or exceed their targets, the QOF has been remarkably successful (Edwards & Langley, 2007), and there is evidence of improved outcomes in areas such as diabetes care (Alshamsan et al, 2010). A large part of this success appears to be that the QOF targets are embedded in clinical practice, with doctors, at least in the early years of the system, coming to regard the QOF measures as being based on best-practice evidence (Maisey et al, 2008).

A key question in relation to the QOF is the extent to which its 'pay for performance' structure, is a central part of the process, or a distraction from it. GPs report that the funding system was not the primary motivating factor, and that instead they were more concerned with meeting the targets because of their basis in clinical practice. This assertion is backed by GPs using the additional funds they received to invest in their practice through employing extra staff or installing better IT systems (Maisey et al, 2008 , McDonald et al, 2007). This claim is also in

accord with research examining the motivation of medical professionals in other settings (Kerpershoek et al, 2016). It is also arguable that the QOF funds do not involve ‘new’ money, but instead the release of funds practices would, had the system not been in place, have received routinely, and so did not provide an additional motivator, but rather one that attempted to achieve behaviour change within the same resource base.

In terms of adverse effects, there seems to be little of the gaming that has plagued PM in hospitals (Checkland et al, 2007), even if GPs have expressed concerns that the QOF is introducing a ‘checklist culture’ into their practice (Hudson, 2009) and that its measures are interfering with the interpersonal contact between doctors and patients (McCartney, 2012), especially in their privileging of biomedical information that routinizes GP work (Checkland et al, 2008). Compared to the considerable problems faced by other PM systems, however, the QOF has been remarkable successful in getting GPs to focus on the outcomes it required of them.

However, it is also the case that the QOF changed over the years, including new targets that GPs regarded as interfering rather than supporting their work, and which attracted criticism from prominent GPs (McCartney, 2016). There is also a more general criticism that the targets for the QOF were set too low in its initial years (Doran et al, 2014), which did have the potential to undermine the credibility of the system, but the extent of behaviour change measured by the QOF suggests did not occur. The Scottish government has now withdrawn the QOF, and its days may well be numbered in England as well. In a similar way to the CPA, the benefits coming from the QOF appear to have occurred most in its early years, with extensions to the system leading to problems which policymakers were unable to overcome.

In all, the QOF, in its early years, motivated GPs and practice nurses to make significant

changes to the organization of their practices, and through that reorganization were able to achieve a great deal of success in meeting the system's targets. There were concerns that inevitably follow the introduction of any PM regime about the targets interfering in autonomous practice, but for the most part the QOF did not appear to have led to the gaming and unintended consequences to the same extent as was present in the PM system introduced elsewhere in the NHS.

### **Performance Management, and Intrinsic and Extrinsic Motivation**

Having examined the CPA and QOF cases, we then looked for, in line with realist methodology (Danermark et al, 2001), theories which offered most potential to add theoretical depth and explanatory power to the cases. We were seeking to explain the link between PM and motivation, with several candidate theories offering themselves. Le Grand's work (1997, 2003) is hugely important in this context, and seeks to generate structures of incentives to support 'knights' (broadly those behaving selflessly) while preventing 'knaves' (those putting self-interest first) from thriving in organizational contexts, as well as allowing service user 'queens' (as in sovereign consumers) to express their choice to the maximum possible extent.

Le Grand's work, however, had drawbacks in that it seems to make assumptions about motivation and agency that are framed within a preference for market-based governance (Greener & Powell, 2009). However, in developing his approach Le Grand (2003) offers the clue to an alternative approach, where he explores intrinsic motivation in the work of Frey and Deci. After comparing the potential explanatory power Le Grand's framework compared to that of Frey and Deci, the latter appeared to offer the most coverage of the cases reported here, so having greater potential to contribute to the social policy and

public management literature. We therefore decided to explore the cases through the lenses of intrinsic motivation theory.

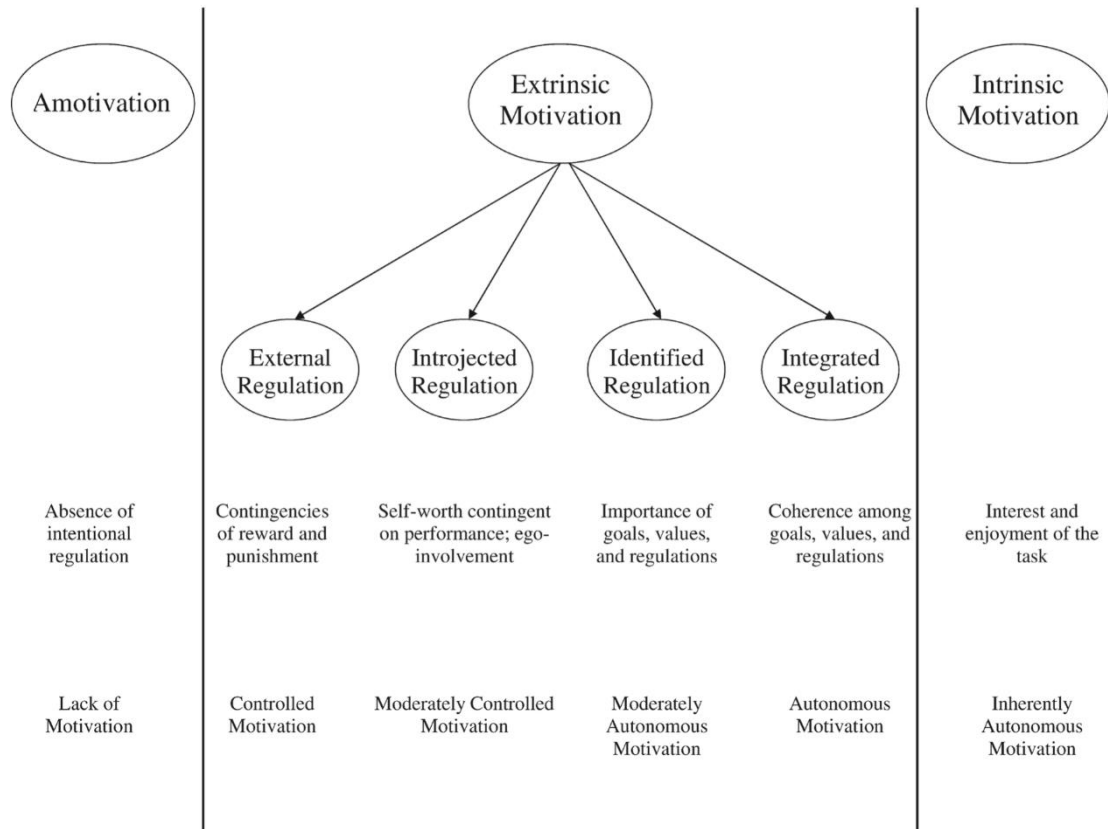
Frey's work on intrinsic and extrinsic motivation is a critique of standard economic theory in that, he suggests, motivation may actually decline in certain circumstances where extrinsic rewards are used to try and change behaviour (Frey, 1994 , Frey & Osterloh, 2001). Frey argues, there is a risk of intrinsic motivation (the desire to do a job for its own sake) being undermined as it is 'crowded-out' if a payment or other extrinsic reward is made instead. Such crowding-out occurs where motivation which was previously seen as discretionary, instead becomes attached to an extrinsic reward and so falls outside of that employee's control. Payment or other extrinsic reward can substitute for the intrinsic desire to do a job well, with the result that above-contract effort is no longer given. This is important because, as Bowles (2016) makes clear, all labour contracts are incomplete as they can never precisely specify exactly what constitutes good work. Employment contracts therefore depend upon employees and employers effectively filling in contract gaps by creating moral or social practices that foster high performance – and a key means of filling these gaps comes through harnessing the intrinsic motivation of public employees. Research suggest that public managers are strongly motivated by public service values (Georgellis et al, 2010) that directly 'crowd-in' behaviours that work toward those ends, provided they are not 'crowded-out' through the careless use of extrinsic rewards instead.

However, there are problems with Frey's framework. There are clear problems in defining 'intrinsic' and 'extrinsic' in a water-tight manner, with the two reflecting a continuum rather than a hard and fast divide, and it can sometimes appear in Frey's work that intrinsic and extrinsic motivation is an either/or, rather than a range of different positions being possible. But the key message coming from Frey's work is that it is crucial to find a balance between intrinsic and extrinsic motivators so that, where possible, those working in organisations are

able to exercise their own locus of control to the maximum extent. This creates the need to maximise intrinsic motivation, while at the same time accepting that the use of extrinsic motivators can also be appropriate and necessary (Jacobsen & Andersen, 2016), especially in cases where tasks may be of little intrinsic interest.

To understand how Frey's framework could be expanded to deal with some of its limitations, the work of Deci (Deci & Ryan, 1985) offers a valuable supplement. Deci (in his work with Gagne, especially (Gagne & Deci, 2005)) expands the intrinsic/extrinsic framework to embrace a wider range of possibilities based on a range of experimental and field studies. At one end of that spectrum is intrinsic motivation where high degrees of autonomous motivation exist, and where interest in the task is high. As intrinsic motivation and task interest gradually fall we see a range of possibilities from where motivation is also autonomous and goals and values are coherently integrated (integrated regulation), through to situations where motivation is externally controlled and self-worth becomes contingent on the achievement of rewards or the avoidance of punishment (external regulation) (for an application to public management, see Chen & Bozeman, 2013). Finally, Deci suggests a final option – amotivation, where there is an absence of intentional regulation and a lack of motivation of any kind. The spectrum of possibilities in Deci's framework are summarised below, with the movement from intrinsic to extrinsic motivation to amotivation moving from right to left.

**Figure one – Gagne and Deci's spectrum of extrinsic/intrinsic motivation possibilities**



Source – Gagne and Deci 2005, p. 336

As such, Deci’s work allows us to consider a much wider range of possible positions between intrinsic and extrinsic motivation, giving the potential for greater theoretical insight than a simple intrinsic/extrinsic distinction.

**Performance Management that works?**

Having outlined a framework for considering the two cases, we can now explore them comparatively, drawing on Deci’s framework to add additional theoretical depth to the analysis.

The CPA’s targets were regarded by local government managers as straightforward and simple to understand (in contrast to ‘Best Value’ which preceded it). Its measures were the result of



consultation between the Audit Commission and local authorities leading to those responsible for implementing them having buy-in. The inspection system had legitimacy because it involved other local authority managers, giving it a strong peer-involvement effect which led to further learning. There were consequences for poor performance for local authority managers who faced censure through both the local media and their professional networks for poor performance, and the threat of electoral difficulties should measured performance be graded in the bottom two levels. Local authorities had autonomy in how they went about achieving their targets, leading to stronger engagement with the targets than might have otherwise been the case. Finally, even though there were problems with local authorities in poorer areas showing the poorest results, this did not appear to undermine the legitimacy of the system as a whole as some councils in this group were able to achieve improvements despite the difficult situations in which they were working.

Viewed from the perspective of Deci's theoretical framework, the CPA is an example of autonomous motivation in that council managers were highly motivated and had integrated their activities in line with their organisation because of their motivation to provide public service. The CPA was not solely based on intrinsic motivators only because of the use of extrinsic rewards (or threats) with the publication of league tables which were often picked up by local media, and were certainly important in professional networks. With autonomous motivation, the design of the performance management system is about working with the desire to provide public service to better integrate the personal values of managers to motivate them to do their jobs well by offering autonomy, but also some degree of extrinsic reward – in the case of the CPA, especially through peer recognition. The success of the CPA can be measured to the extent that it achieved this mix of different motivating factors, as well as its appearing to led to increases in measured performance, and this appeared to be borne out in its results.

The CPA offered broad targets, which required ‘translation’ into more everyday activities that were relevant to council managers, which required managers to bridge the gap between high level, general targets and the everyday practices of public organizations if they are motivated to do so. Managers translated the targets from abstract levels into a form which could motivate and engage those responsible for meeting them. Poor performers in the CPA faced no financial penalty, but potentially the threat of peer, media and electoral censure which might have limited their professional careers. For high performers, the promised additional funding and freedoms did not emerge, suggesting that the extrinsic rewards of recognition (or sanction) through the media and professional networks were sufficient, even overcoming the disappointment that must have resulted.

The QOF’s targets were also the result of extensive consultation between central government and health care professionals, and so were seen as holding legitimacy which was then increased by the targets being presented to GPs as being evidence-based. The increased funding for practices coming from the QOF led to GPs investing in information systems to allow them to better monitor their own levels of performance with the meeting of targets becoming an everyday practice through the incorporation of ‘cues’ into GP consultations. The nature of GP-work leads us to expect intrinsic motivation to be high because of its strong public service element, but also because of the long training required to qualify, and which require significant motivation to achieve.

The reduction in autonomy that the QOF targets involved, combined with the use of explicit extrinsic motivators suggests that the QOF was an example of identified regulation rather than the integrated regulation of the CPA. The task of performance management systems in identified regulation is to maintain high levels of intrinsic motivation by keeping GPs interested in their roles, but also continuing to allow them substantial autonomy, while also accepting that

extrinsic motivators are present. The complexity of the balancing act between intrinsic and extrinsic motivation involved in identified regulation brings a range of risks. The QOF's routinisation of clinical work meant that such autonomy is increasingly only in terms of GPs ability to manage their surgeries, and this may not be what many doctors entered the medical profession to achieve.

The QOF offered GPs intrinsic rewards in terms of knowing they are doing a good job by achieving their targets (with those targets associated with quality clinical practice), but also a sliding scale of funding commensurate with their achievement against their targets, and should the GPs have come to regard the extrinsic motivators as not supporting their work, there is was risk of their motivation being crowded-out. Poor performers in the QOF faced no professional threat of losing their job, but the sliding scale of pay for performance, and the association of the QOF with good practice meant that there were real professional risks attached to reduced funding. Equally, should the link between the measures and clinical evidence be broken (which some GPs did claim (McCartney, 2016)), this would risk a crisis of legitimacy very quickly as doctors would perceive them to be controlling rather than supportive of their work. Finally, there was a danger of status professionals having their intrinsic motivation being crowded-out through the extrinsic rewards, especially those of the pay-for-performance system, which could substitute for their own motivation to do the job for its own sake, and lead to what Deci calls 'over-motivation', and intrinsic motivation being crowded-out.

The CPA and the QOF therefore provided a mix of intrinsic and extrinsic incentives. The CPA, because it was seen as legitimate by council managers, meant that those who achieved high ratings were given respect and status within local government networks from their peers – and this form of respect, based on sustained achievement rather than meeting specific, more narrow goals, was a key mechanism for further internalising organisational goals and enhancing

intrinsic motivation further (see Frey, 1994, for an examination of the importance of this). The QOF had a reward structure that was more explicitly based on extrinsic rewards because of the payments available for meeting and exceeding performance targets. However, the QOF reward system was based not on individual performance but on that achieved by GP practices, with resources going to the practices rather than to individual doctors or managers within them. In the QOF, collective performance in line with best practice was rewarded, so that intrinsic motivation was generally not substituted for extrinsic – the extrinsic motivators led to an increased internalisation of the targets and for managers to be highly motivated in achieving them. Through this lens, the QOF represented less ‘pay for performance’, and more an extrinsic reward system that supported already-present intrinsic motivation to provide public service. It succeeded in the 2000s because the system supported the already existing intrinsic motivation of the staff to do a good job.

Both the CPA and the QOF allowed local autonomy in how their targets were to be achieved, and this was crucial in achieving buy-in and legitimacy for the targets, and for professionals to retain their locus of control. The CPA set its targets at a level abstracted from everyday practice, and so local authority managers needed to find ways of specifying lower levels goals and organizing services to achieve the higher level targets, but allowing considerable local discretion in how they were to be achieved. QOF targets, in contrast, were highly specified and so initially might appear to offer less scope for flexibility in how they could be achieved, but practices were innovative in how they met the challenge, redrawing professional boundaries so that GPs and practice nurses could focus on cases more suited to their expertise, and finding new approaches to teamworking to reach or exceed their targets. What both cases share is that the performance system specified what needed to be achieved, but not how, and this allowed GPs or local government managers on the ground to organise their work, giving them ownership of the new systems and allowing them to experiment and innovate in result.

## **Conclusion – Learning from the CPA and the QOF**

It is worth emphasising again that this paper is by no means claiming that the CPA and the QOF were perfect, but rather they have elements of success that make them rare exceptions where some measure of success in performance management has been achieved, and so that offer us clues as to the principles through which we might improve the design of such systems. Equally, we are not arguing that PM is the only thing that matters in motivating staff. The cases presented above and the subsequent discussion do, however, present important learning concerning the introduction and implementation of PM systems in the public sector, as well as giving an illustration of the power of Deci's framework in its application, both of which we believe to be original contributions to the literature.

First, in the design of a PM system, it is crucial to get those that will lead the process of implementation engaged and supportive of the system. In Frey's and Deci's terms, the locus of control must be with those that are responsible for driving performance improvement, who must believe that the PM system is there to support rather than control them. Both the CPA and the QOF systems involved consultation between those designing the system and the key professionals who would oversee their success, and so upon whom successful implementation ultimately depended. This increased the perceived legitimacy of the measures chosen for the system. That policymakers made the effort to establish the need for improvement and then engaged in consultation with key professional groups so that those who would be responsible for implementing the systems accepted them to a large degree. Without that effort the PM systems would have been much harder to subsequently implement successfully.

Second, getting engagement with the PM systems from public professionals with strong

intrinsic motivation required them to be seen not as politically-driven, extrinsic targets, and instead regarded as being legitimate and so internalised into decision-making by those professionals. In the cases above, the targets were initially set at very different levels – in the CPA they were set at a level abstracted from the everyday practices of council workers, whereas the QOF targets were more specific and embedded in everyday healthcare practices. This meant that local government managers were successful in both translating the abstract targets into everyday practices as well as internally ‘selling’ the targets as being important to their staff so that the targets were internalised and regarded as intrinsic to the performance of their organisations. Without this selling role, there was a danger the abstract, high-level targets would be regarded as being imposed for extrinsic reasons, and so a source of gaming, as they did in hospitals in England. The QOF’s targets seem to have avoided gaming because they were perceived as legitimate and embedded in evidence-based practice, appealing to GP’s intrinsic motivation and so leading to practice teams being prepared to put in the work to achieve them.

Third, despite the importance of intrinsic motivation in public service, there is the importance of there being a link between measured performance and reward (or penalty) for the staff responsible for achieving the targets. It would be easy to express this in terms of simply getting incentives right, but this language over-simplifies the problem, and runs the risk of suggesting that staff are motivated entirely by extrinsic factors. A key lesson from the use of PM in the CPA and the QOF is that a balance of extrinsic and intrinsic factors is crucial in getting PM to work. Extrinsic factors in the CPA worked in the form of public reward or censure through peers, the media and the general public through elections. The promised additional freedoms and sources of funding did not emerge for high performers. If we take a wholly rationalistic view, we might have expected this would result in only low performers, or councils that were concerned about falling into the bottom two classifications of performance, being concerned

about the impact of PM as there was little extrinsically to push them to try and achieve high performance. But this was not the case. The reason for this has an extrinsic element in the high performing managers received recognition in local government networks for their achievement, perhaps enhancing their future job prospects as a result. The CPA appears to have provided a framework to achieve a focus for service improvement despite there being little in the way of direct reward for local authority managers.

In the QOF, a different sets of motivators was in place. Here there were financial penalties and rewards for reaching threshold levels on the indicators and targets, leading to the system sometimes being regarded as 'pay-for-performance' in the literature (Grant et al, 2009). Classifying the QOF system in this way can lead to us interpreting it as an extrinsically-driven one. It is certainly the case that the extra resources high performance attracted were important to GPs. But at the same time, GPs suggested that they did not used additional performance funds to reward themselves, but instead invested in their practices by hiring extra staff or improving information systems. This infrastructural investment suggests that GPs were prepared to invest in the achievement of the QOF targets, and so were either behaving extremely cannily and strategically to boost their long term returns from the system, or that they regarded its targets as being legitimate endeavours for them to try and improve against (or perhaps both). Because GPs regarded the QOF targets as representing best practice, this provided them with the intrinsic motivation to work to improve them (and to work long hours on top of their clinical practice to that end), as well as reorganising their practices, working more closely with other health professionals, and using QOF funds to invest in computer systems to support future work. This suggests the QOF successfully utilised a mix of intrinsic and extrinsic motivating factors.

If professionals are to be performance managed, it is crucial to respect their autonomy and

allow them to design the systems for the highest possible level of attainment rather than policymakers attempting to specify not only what the targets are, but how they will be achieved. In both the CPA and the QOF, as well as allowing professional groups to be involved in the design of the systems, professionals were allowed discretion ‘on the ground’ in how they actually went about achieving the targets. This again provided intrinsic motivation, with professionals being given the autonomy to organise their work themselves, even though they knew that the results of their work would be closely measured. In healthcare more generally it seems that successive governments not only want to impose top-down performance measures but also to specify how local health services are to be organised. This central interference not only in setting what the goals of public services should be, but also in how those targets are to be achieved, risks alienating professionals who are left with less decision space (Peckham et al, 2008) to seek improvements and to exercise their professional judgement – key factors again in achieving the levels of intrinsic motivation that are necessary for sustained improvement.

The two cases also highlight a key problem with PM systems. Both the CPA and the QOF appear to have approached the limits of what they could achieve, even despite the success at least initially achieved by them. This raises the question of whether PM systems, no matter how well initially designed, have shelf-lives, and what we need to do when PM systems no longer deliver improvements. However, considering the findings above, in those circumstances it becomes necessary to begin a new round of consultation and discussion with public professionals, preferably based around evidence, in order to establish the basis of a new system that might be nuanced enough to detect further improvements, or which might incorporate new areas where improvements are necessary. All PM systems have their limits, but following the principles outlined above gives a way forward when their current designs run up against their limits.



The approach outlined here address many of the problems with PM systems identified at the beginning of the paper. The challenge of measuring outputs is explored in terms of them being developed in a more nuanced and discursive manner, and this to some extent also reduces the risk of measures not capturing the correct underlying phenomenon. The evidence presented above also suggests that managers are less likely to manipulate targets they have had a role in setting, and that those targets can address the long-term goals of the organisation and not simply be set around short-term expediency. However, setting the correct levels of targets will always be an on-going challenge – and the key appears to dealing with that comes from regarding targets as a part of a PM system, alongside ‘soft’ information, discussion and negotiation (Goddard et al, 1999), rather than the over-simplistic approach that is often taken to them at present. PM systems are always subject to potential abuse, and cannot deal with the problem of politicians imposing their own views upon public organizations, which also appears to be something of an intractable problem.

Despite these problems and the poor reputation it has achieved in the 2000s, PM can work better than most of its current implementations. But it requires carefully thinking about the principles of design in the systems to incorporate relevant professionals, for the design to be seen as tackling legitimate problems, and for a mix of intrinsic and extrinsic rewards to be available so that public managers believe they are not only doing a good job, but are seen to be doing a good job (because the system measures it as such). With careful use, and as one of many public management tools, PM can be a central part of the process by which public services are improved. It is testament to the lack of care which policymakers have treated it so far that its use has fallen into such disrepute.

## REFERENCES

- Alshamsan R, Millett C, Majeed A, Khunti K. (2010). Has pay for performance improved the management of diabetes in the United Kingdom? *Primary Care Diabetes*, 4, 73-8
- Andrews R, Boyne G. (2012). Structural Change and Public Service Performance: The Impact of the Reorganization Process In English Local Government. *Public Administration*, 90, 297-312
- Andrews R, Boyne G, Law J, Walker R. (2005). External Constraints on Local Service Standards: The Case of Comprehensive Performance Assessment in English Local Government. *Public Administration*, 83, 639-56
- Bertelli A, John P. (2010). Government checking government: how performance measures expand distributive politics. *Journal of Politics*, 72, 1-14
- Bevan G, Hood C. (2006). What's Measured Is What Matters: Targets And Gaming In The English Public Health Care System. *Public Administration*, 84, 517 - 38
- Bloomfield B. (1991). The Role of Information Systems in the NHS: Action at a Distance and the Fetish of Calculation. *Social Studies of Science*, 21, 701-34
- Bowles S. (2016). *The Moral Economy: Why Good Incentives Are No Substitute for Good Citizens*, London: Yale University Press
- Boyne G. (2002). Concepts and Indicators of Local Authority Performance: An Evaluation of the Statutory Frameworks of England and Wales. *Public Money and Management*, 22, 17-24
- Boyne GA, James O, John P, Petrovsky N. (2012). What if Public Management Reform Actually Works? The Paradoxical Success of Performance Management in English Local Government. In *Paradoxes of Modernization*, ed. H Margetts, P 6, C Hood, pp. 203-18. Oxford: Oxford University Press

- Carter N, Klein R, Day P. (1992). *How organisations measure success: the use of performance indicators in government*, London: Routledge
- Chapman R, Dunsire A, eds. (1971). *Style in Administration: Readings in British Public Administration*. London: George Allen and Unwin
- Checkland K, Harrison S, McDonald R, Grant S, Campbell S, Guthrie B. (2008). Biomedicine, holism and general medical practice: responses to the 2004 General Practitioner contract. *Sociology of Health & Illness*, 30, 788-803
- Checkland K, McDonald R, Harrison S. (2007). Ticking Boxes and Changing the Social World: Data Collection and the New UK General Practice Contract. *Social Policy & Administration*, 41, 693-710
- Chen C, Bozeman B. (2013). Understanding Public and Nonprofit Managers' Motivation Through the Lens of Self-Determination Theory. *Public Management Review*, 15, 584-607
- Cowell R, Downe J, Martin S, Chen A. (2012). Public confidence and public services: it matters what you measure *Policy & Politics*, 40, 123-43
- Danermark B, Ekstrom M, Jacobsen L, Karlsson J. (2001). *Explaining Society: An Introduction to Critical Realism in the Social Sciences*, London: Routledge
- Deci E, Ryan R. (1985). *Intrinsic motivation and self-determination theory in human behaviour*, Plenum: New York
- Doran T, Kontopantelis E, Reeves D, Sutton M, Ryan A. (2014). Setting performance targets in pay for performance programmes: what can we learn from QOF? *British Medical Journal*, 348, 1595-602
- Dunleavy P, Hood C. (1994). From Old Public Administration to New Public Management. *Public Policy and Management*, July-September, 9-16

- Edwards A, Langley A. (2007). Understanding how general practices addressed the Quality and Outcomes Framework of the 2003 General Medical Services contract in the UK: a qualitative study of the effects on quality and team working of different approaches used. *Quality in Primary Care*, 15, 265-75
- Exworthy M, Wilkinson EK, McColl A, Moore M, Roderick P, et al. (2003). The role of performance indicators in changing the autonomy of the general practice profession in the UK. *Social Science & Medicine*, 56, 1493-504
- Flyvbjerg B. (2001). *Making Social Science Matter: Why social inquiry fails and how it can succeed again*, Cambridge: Cambridge University Press
- Frey B. (1994). How Intrinsic Motivation is Crowded Out and In. *Rationality and Society*, 6, 334-52
- Frey B, Osterloh M. (2001). *Successful management by motivation: Balancing intrinsic and extrinsic incentives*, London: Springer
- Gagne M, Deci E. (2005). Self-determination theory and work motivation. *Journal of Organizational Behaviour*, 26, 331-62
- Georgellis Y, Iossa E, Tabvuma V. (2010). Crowding Out Intrinsic Motivation in the Public Sector. *Journal of Public Administration Research & Theory*, 21, 473-93
- Goddard M, Mannion R, Smith P. (1999). Assessing the performance of NHS Hospital Trusts: the role of 'hard' and 'soft' information. *Health Policy*, 48, 119-34
- Grant S, Huby G, Watkins F, Checkland K, McDonald R, et al. (2009). The impact of pay-for-performance on professional boundaries in UK general practice: an ethnographic study. *Sociology of Health & Illness*, 31, 229 – 45
- Greener I, Powell M. (2009). The Other Le Grand? Evaluating the 'Other Invisible Hand' in Welfare Services in England. *Social Policy & Administration*, 43, 557-70
- Gunstone C. (2007). QOF. *British Journal of General Practice*, September, 748

- Haubrich D, Guitierrez-Romero R, McLean L. (2008). Performance and Deprivation in English Local Government: Correlations and Possible Implications. In *Oxford Policy Institute 'Institutions, Incentives and Public Sector Performance'*. Oxford
- Heath G, Radcliffe J. (2007). Performance Measurement and the English Ambulance Service. *Public Money & Management*, 27, 223-8
- Hudson S. (2009). Is the QOF Failing to Motivate GPs or Improve Patient Care? Evidence from an Exploratory Study. *Journal of Primary Care*, September, 1-5
- Jacobsen C, Andersen L. (2016). Leading public service organizations: how to obtain high employee self-efficacy and organizational performance. *Public Management Review*, Online early article, 1-21
- Kerpershoek E, Groenleer M, de Bruijn H. (2016). Unintended responses to performance management in dutch hospital care: Bringing together the managerial and professional perspectives. *Public Management Review*, 18, 417-36
- Klein R. (1982). Performance, Evaluation and the NHS: A Case Study in Conceptual Perplexity and Organizational Complexity. *Public Administration*, Winter 60, 385-407
- Le Grand J. (1997). Knights, Knaves or Pawns? Human Behaviour and Social Policy. *Journal of Social Policy*, 26, 149-69
- Le Grand J. (2003). *Motivation, Agency and Public Policy: Of Knights, Knaves, Pawns and Queens*, Oxford: Oxford University Press
- Lowe T, Wilson R. (2015). Playing the Game of Outcomes-based Performance Management: Is Gamesmanship Inevitable? Evidence from Theory and Practice. *Social Policy and Administration*, On-line early
- Maisey S, Steel N, Marsh R, Gillam S, Fleetcroft R, Howe A. (2008). Effects of payment for performance in primary care: qualitative interview study. *Journal of Health Services Research and Policy*, 13, 133-9

- Mannion R, Goddard M, Smith P. (2001). On the Limitations and Pitfalls of Performance Measurement Systems in Health Care. In *Quality in Healthcare: Strategic Issues in Health Care Management*, ed. H Davies, M Tavakoli, M Malek, pp. 158-69. Aldershot: Ashgate
- McCartney M. (2012). *The Patient Paradox: Why Sexed Up Medicine is Bad for your Health*, London: Pinter and Martin Ltd
- McCartney M. (2016). *The State of Medicine: Keeping the Promise of the NHS*, London: Pinter and Martin
- McDonald R, Checkland K, Harrison S. (2009). The new GP contract in English primary health care: an ethnographic study. *International Journal of Public Sector Management*, 22, 21-34
- McDonald R, Harrison S, Checkland K, Campbell SM, Roland M. (2007). Impact of financial incentives on clinical autonomy and internal motivation in primary care: ethnographic study. *BMJ*, 334, 1357
- Mclean I, Haubrich D, Gutierrez-Romano R. (2007). The Perils and Pitfalls of Performance Management: The CPA Regime for Local Authorities in England. *Public Money and Management*, 27, 111-8
- Osborne D, Gaebler T. (1993). *Reinventing government*, New York: Addison Wesley
- Pawson R. (2006). *Evidence-based Policy: A Realist Perspective*, London: Sage
- Pawson R, Greenhaigh T, Harvey G, Walshe K. (2005). Realist Review - a new method of systematic review for complex policy interventions. *Journal of Health Services Research and Policy*, July 10, 21-34
- Pollitt C. (2013). The Logics of Performance Management. *Evaluation*, 19, 346-63
- Pollitt C, Bouckaert G. (2011). *Public Management Reform: A Comparative Analysis*, Oxford: Oxford University Press

Tan H, Harvey G. (2015). Unpacking the Black Box: A realist evaluation of performance management for social services. *Public Management Review*, Online early, 1-23

van der Geer E, van Tuijl HFJM, Rutte CG. (2009). Performance management in healthcare: Performance indicator development, task uncertainty, and types of performance indicators. *Social Science & Medicine*, 69, 1523-30