Written submission to Public Administration Select Committee's inquiry into public sector performance targets

John Seddon MSc (dist), BSc (hons), FIQA.

January 2003

Summary of evidence

Question 31 of your brief asks:

"If you believe the use of targets is a bad or flawed idea, what alternative approach would you advocate which would help bring about real and lasting public service improvements?"

This evidence argues that targets are inherently flawed. I seek to demonstrate – with case studies – that targets actually undermine achievement of purpose, which is improving public sector performance. I shall also illustrate an alternative approach, which is to make capability measurement the corner stone of public sector improvement. I shall describe what these measures are, how they differ from targets and how they are used to understand and improve performance in a sustainable way. The case studies show significant performance improvement using capability measures and I shall ask the reader to reflect on whether this level of improvement could ever have been achieved within the current targets regime.

The case studies are illustrative of the general problems; the arguments that apply in these cases apply to every public sector example I have knowledge of.

The evidence is presented in the following structure:

- Contributor's background
- Comments on your evidence to date
- Case studies
- Conclusions
- The future of the specifications and inspection regime

Contributor's background

I am an occupational psychologist, consultant and management thinker. I am managing director of Vanguard Consulting. In the early Eighties I was researching the reasons for failures of major change programmes. Based on what I learned I have developed a more effective method of change for performance improvement. My work is a combination of systems thinking - how work works - and intervention theory - how to change it. I credit W. Edwards Deming for teaching me what is wrong with conventional ('command and control') management thinking and I credit Taiichi Ohno (who built the Toyota Production System) with introducing me to the practices and principles of systems thinking as applied to operational performance. I have specialised in translating these ideas for service organisations.

I have been a leading critic of management standards and models, in particular ISO 9000; which is, quite simply, based on bad theory. In my view it is management thinking that needs to change. ‘Command and control' is a failing management paradigm and I propose instead that managers learn to adopt a systems perspective. It is a better way to design and manage work.

In the last two years I (with the Vanguard team) have been invited to work in the public sector. I have learned there is appalling waste and poor morale. Both are caused by the specification and inspection regime whose purpose, paradoxically, is to improve public sector performance; targets are the corner stone of this regime.

Comments on your evidence to date

I have read the evidence you have received to date. Much of the evidence might lead you conclude that targets are of value because:

- They give direction
- Without them there would be disorder

While a target if expressed in general terms may give clarity of direction, a numerical target is more likely to increase disorder in the system – the way the work works. I shall illustrate this phenomenon with examples.

There has been general acknowledgement of the risks associated with targets. To have too many, it is argued, is counter-productive. But if it is true, as I maintain, that the inherent nature of a target is flawed, to have less is not an improvement. It is the nature of measurement that needs to change.

The general advice is that targets can be well set. Nobody offers an unequivocal view of how to achieve that; they cannot for there is no reliable method. Many offer the opinion that the involvement of the recipient is of crucial value. While this is an admirable sentiment, it is not a solution if the concept of a target is flawed.

You have heard it said that it is normal for one not to achieve all of one’s targets. This sits uncomfortably with another idea: stretch targets. In truth, targets are arbitrary measures; hence both of the foregoing ideas might be valid and the fact of that exposes the inherent problem. Targets are flawed; they do not pass the test of a good measure:

Can this help us understand and improve performance?

There have been many observations on the value and extent of the specification / inspection / assessment / monitoring regime. In recent months I have noticed a softening of the intensity with which inspection is administered. I have no doubt it reflects a realisation that things are not working, but there has not been sufficient questioning as to whether the regime itself is a cause of the failure to improve public sector performance. Not only does the regime employ flawed measures, it has created a massive bureaucracy to feed its purpose. This constitutes an enormous amount of waste. Rather than being enthused with a sense of contribution to improvement,

public sector personnel find themselves being evaluated on measures that they often do not perceive to be relevant to their task in hand and burdened by reporting and administrative procedures that add no value to their work. The consequential impact on morale amongst managers and workers alike is something that should concern us deeply.

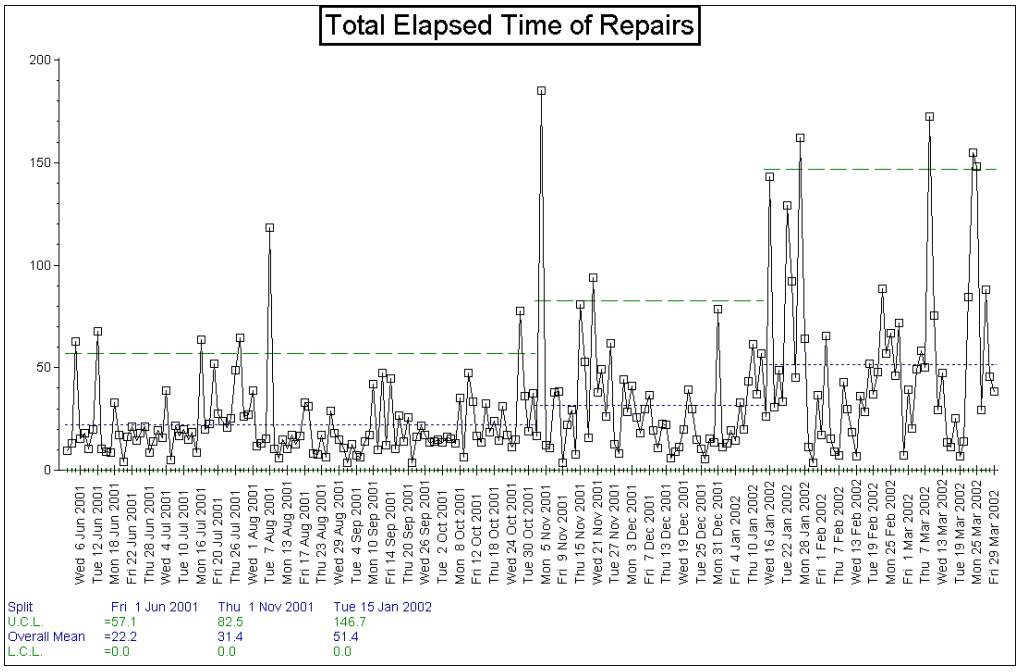
There has been general acknowledgement that targets can drive the wrong behaviour. There is some disagreement as to the extent of this problem. This is typical of the correspondence I receive:

"My wife has just spent part of the weekend moving around some patients' appointment dates, to meet a target of 21 weeks max wait. This creates extra work for the clerks and the patients, for no benefit. Happening throughout the NHS."

The suggested response to the phenomenon has been to 'police' the 'cheats'. This is to compound the problem, adding complexity and cost. Targets drive the wrong behaviour in ways that are endemic, systemic and ubiquitous. To deny this is either to be out of touch with where the work is done or in fear of the consequences of being open about it.

Case studies

I refer first to the housing repairs case discussed in the articles that accompany this evidence. This organisation was subject to BVPI targets and was achieving those targets. To establish a capability measure one asks: what is the purpose in customer terms? The answer is to make repairs and how long this takes – end-to-end time – is what matters to tenants. Measuring end-to-end time revealed the following:



Capability measurement shows a picture that cannot be 'seen' with BVPI data. From the tenants' point of view the average time it takes to complete a repair in March 2002 is 51 days. It could take as long as 146 days and as little as 1 day. The capability measure tells us that both 1 and 146 days are just-as-probable results; the system is very unstable. How then can this organisation be achieving its BVPIs?

Firstly, targets were being achieved through 'cheating'. Jobs were closed and re-opened even though they had not been completed, sometimes with 'justification' – ' if tenants are out this should not count'. Secondly, 'cheating' occurs with changing job classifications to meet times – is this 'an emergency', 'urgent' and so on. You may think these people should be held to account for their behaviour but we have found this phenomenon in every case in our public sector work. Thirdly, one repair from a customer's point of view may be four jobs in this system. To repair a window may require four trades, each would have a job sheet and each of these would be subject to the BVPI regime. The purpose of the system is to comply with targets, thus the modus operandi is 'open and close jobs' not repair properties. Peoples' ingenuity is focused on the wrong things.

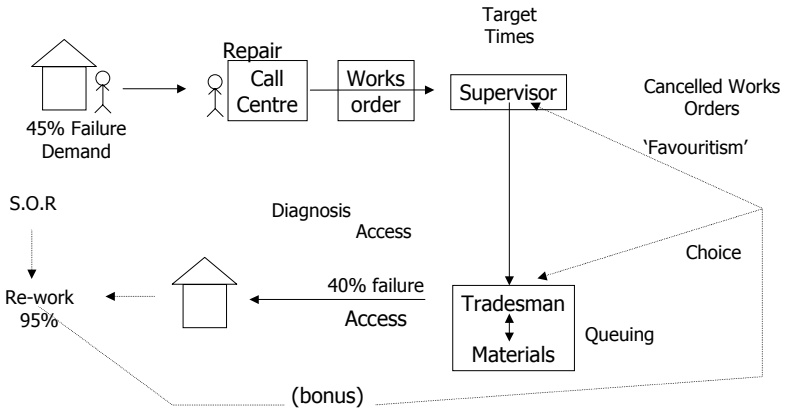
A capability measure invites one to question the causes of variation. In the chart above variation increased significantly on two occasions. In October 2001 a new management structure was put in place. New supervisors, keen to be the best in terms of achieving BVPIs, actually de-stabilised the system (yet they were unaware[1]). In November 2001 a call centre was introduced, something mandated by government policy. This caused further de-stabilisation. Again, no one knew until the capability measure invited the question.

The capability measure tells us about the 'what' of performance – how well the system is achieving its purpose. The next step in performance improvement is to find out 'why', to understand the causes of variation; the things, in this case, that make time go longer. While we have identified the two major causes, there are more.

Here is a system picture of the repairs organisation:

[1] You may like to reflect on the extent of this unintended consequence of target setting.

Housing repairs



Some notes of explanation:

Approximately forty percent[2] of demands into the call centre were 'failure demands' – demands caused by a failure to do something or do something right for the customer – for example tenants progress-chasing their repair or complaining that the repair had not been completed to their satisfaction. The remainder were 'value demands' – people requesting repairs to their properties. The failure demands clogged the system – the call centre workers would have to problem-solve these and get back to the customer with an answer. It often took time to locate tradesmen or supervisors to get an answer for the tenant.

The call centre worker is effectively responsible for diagnosing the reported problem and determining its solution – that is to say determining the work to be done. This was given a specification using what is called the Schedule of Rates; this in turn will determine how the tradesman is paid. Tradesmen would dispute the work specified on the Schedule of Rates most of the time (in fact in excess of 90% of the time). Because of this an administrative function, a 'cottage industry', had been established to deal with these matters. The administrators would take returned works orders from the tradesmen upon which the tradesmen had altered the Schedule of Rates code and pass the same to the tradesmen's supervisors who would make a judgement as to what was correct. Subsequent changes would need to be returned for further administration. None of this adds any value to doing the work – it is all waste.

Arranging access is also done in the call centre. Yet supervisors would allocate work according to value (earnings) to the tradesmen – and favouritism could play a part in allocating work. Tradesmen would schedule their work to maximise their earnings. As a consequence tradesmen often had problems with gaining access and performing the repair. In addition the tradesmen would have to wait for up to an hour each morning, queuing to get their materials.

[2] In other cases failure demand has been found to run as high as eighty percent.

All of these problems had been created by design. Managers may believe that this organisation would work just fine if everybody 'did as they should'. But such thinking ignores the fact of variety. To 'command and control' service delivery is as much a problem for these managers as it is for Government. To design a service that works one needs to learn how to design against demand, to understand the nature and extent of variety in demand and optimise the way the system works in response to that.

Diagnosing a repair could never be satisfactorily achieved by two parties – the tenant and the call centre worker – who know little about the expertise of the tradesmen. Turning this diagnosis into a specification and linking that to pay are the conditions that lie at the heart of the system's failure. The waste in this system included: re-visiting the properties, re-working the Schedule of Rates paperwork, disputes with respect to pay, doing more than was required in the repair hence wastage of materials and labour, and so on.

Having gained knowledge about the 'what and why' of current performance, the people who did this work re-designed it. The first step in re-design is to clarify the value work. In this case it can be described as: diagnosis, access and repair. The re-design was as follows[3]: The customer called the call centre who routed the call directly to a tradesman who was working on the estate (they had learned that demand was predictable by geography and thus had determined where to be located). The tradesman would arrange to visit the tenant by mutual agreement. The tradesman would then arrive and, if possible, complete the repair (understanding demand had led them to learn what materials to carry). If it were not possible to complete the repair, for reasons of materials or trade skills, the tradesman would arrange for a repair at an agreed date. Within weeks the end-to-end time for repairs fell. All repairs were being completed in eight days. As well as transforming performance, the change transformed morale.

It would be wrong to assume the solution developed in this case should be made a national prescription. It would be a mistake to prescribe method. The prescription required is higher level; it is simply to use measures that invite thoughtful questioning about how the work works. We need to liberate method, not constrain it.

The consequence of targets was an increase in disorder in the system. This was in part because the targets drove the wrong behaviour but also because the targets and their associated activity bore no meaningful relation to the work.

By contrast, capability measures increase order in the system by helping people understand and work on the causes of variation. Capability measures drive more productive behaviour. Capability measures are not arbitrary – they are actual; and because they are derived from the work they provide value to those who do the work. In short they pass the test of a good measure.

---

[3] This is the solution adopted by this example. In other examples of housing repairs other variations of this solution have been developed. The important observation is that method can and should vary according to local circumstances – design is against demand. The thing that remains constant to all is measurement against purpose.

Before we move on from the housing sector, I want to comment on the target that 70% of maintenance work should be planned. The consequence is higher costs but those who set the target are unaware of this. Often repairs have to wait to be completed in the planned programme, causing distress and sometimes further problems. Frequently serviceable items are replaced unnecessarily because they are in the plan. Someone somewhere thinks a planned world is better than a reactive world. If only this person knew the costs of thinking this way.

The housing repairs case illustrates the impact of numerical targets on the behaviour and performance of the system. Government also promulgates targets that specify method – how work will be done. For example, the target to have "100% of legally permissible services e-enabled by 2005" (BV 157) should concern us all. It is a specification driven by opinion not knowledge. The recent report showing a lack of interest in and use of Government web sites should cause us to question our assumptions about the nature of citizen demand. Much of the demand into Local Government call centres is what I describe as 'failure demand' – demand caused by a failure to do something or do something right for the customer. The Government's strategy will result in the institutionalisation of waste by moving failure demand to a call centre and treating this as 'normal' work. The consequences will be high costs and poor services and such a re-structuring of work will create a barrier to future improvement of the services. There is a better way to solve this problem that improves service and reduces costs; the flaw remains that those Authorities that take this route will fail to meet the Government target.

The e-enabled access target is a general specification that will result in the purchase of information technology and telephony equipment at a significant capital cost that may be used in a variety of ways for good or ill. Government also uses targets in the sense of detailed activity specifications. Benefits processing is a good example of a detailed specification dictating work methods. The Department of Work and Pensions has promulgated manuals detailing the requirements to manage benefits. It may be a well-intended intervention, but it is failing its purpose. Gordon Brown insists on no investment without reform. He is investing in excess of £200M in implementing centrally specified alterations to the way benefits processing is managed. I am confident he will not get a return.

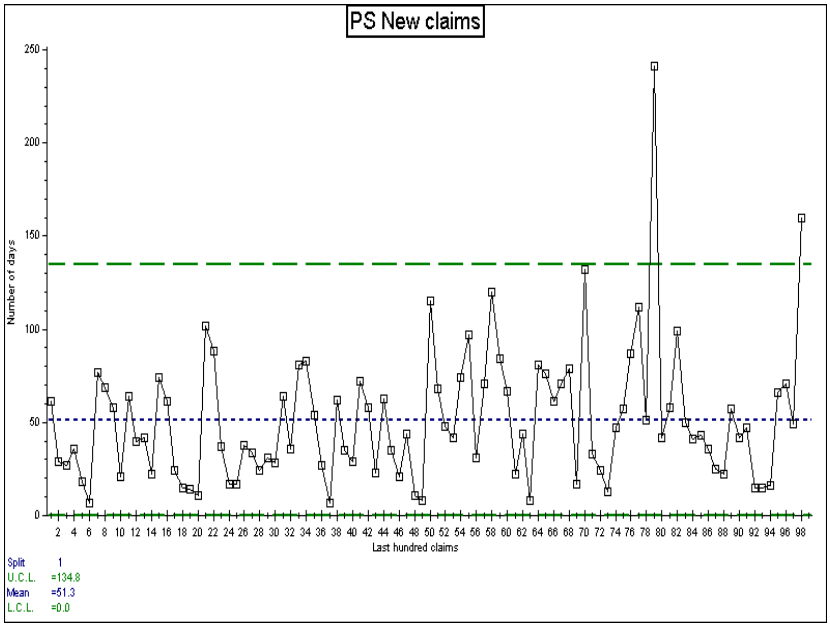
Benefits processing consists of a front office, where claimants are dealt with, and a back office, where the benefits are calculated and paid. The two are usually connected by electronic means – a document image processor – documents are scanned and held on a central database. As is the case with all specifications, those who write them think of things they can measure which seem consistent with doing things properly. While there are a host of standards and targets in the DWP specification (all of which need establishing and monitoring, creating a bureaucracy) here are the essential few that, paradoxically, sub-optimise the system:

Front office: Time to see claimants, time to respond to correspondence
Back office: Time to determine the benefit and pay it

I have little doubt the reader would be thinking 'why?' for these seem like quite reasonable things to focus on. Yet as with the housing repairs example, these measures actually create disorder, they de-stabilise the system.

When Local Authority personnel study benefits processing as a system they learn that there is a high level of failure demand in the front office, people progress-chasing and, more importantly, people not coming in with everything that is required to determine their claim. People in the front office send what they have for scanning, to meet service requirements, and ask the applicant to return with whatever else is required. Document image processors – 'scanners' – require that work is sorted and batched into like work types. This means applicants' information is separated and thus needs to be re-collated electronically. Inevitably documents are poorly scanned, duplicated, lost or wrongly sorted; applicants are frequently asked to bring in things they have already provided. In the back office the clock for the performance measure only begins when all of the required information is to hand. Achieving this is hampered by the way work is designed and managed. It is relevant to note the DWP specification encourages the use of document image processors.

Opening up these problems starts with looking at the end-to-end time for processing benefits from the applicants' point of view, establishing a measure of capability. Here is the capability chart for one case:



The capability measures shows it can take anything up to 134 days to process a benefit from the applicants' point of view. Anything from 1 day to 134 days are just-as-probable events. As with housing repairs the causes of variation are in the way the work is designed and managed. In this case the underlying cause is the DWP specification, it is creating disorder in benefits processing.

The Local Authority whose data is reported above has re-designed benefits payment processing using systems principles, removing all major causes of variation. They currently process all benefits in 8 days. The national average, I am told, is 60 days.

Nothing written in the DWP specification would have aided these people in making this change; following the specification obviates improvement.

Conclusions

Can a target can be well set? How could targets for the examples above have been well set? Should we base a target on current performance or national comparisons? Should targets be set to encourage the same level of performance, better performance or even 'stretched' performance? How would we determine the right number? Would we get the right answer if the recipient were involved? Could anyone explain how a target might have been set that would have predicted the improvement achieved in these examples? If the DWP set a target to process all claims in eight days people would protest and/or demand more resources. If social landlords were told they are to repair all properties in eight days, would people be motivated to achieve it or respond with incredulity?

Targets are arbitrary, capability measures are not – they are derived from the work, not plucked from the air. Targets increase disorder in systems, capability measures lead people to act in ways that increase order and control. Targets focus people on the wrong things, they focus peoples' ingenuity on survival not improvement. Capability measures encourage peoples' ingenuity to be focused on how the work works. Targets have no value in understanding and improving performance; capability measures are of immense value in understanding and improving the work. Targets de-motivate, capability measures motivate, because they put control and understanding in the right place.

Setting a target well boils down to one argument, you need to have the appropriate experience. As the cases above have shown experience would mislead, any targets based on experience would almost certainly maintain the status quo. Deming would say: don't rely on experience; it is no substitute for knowledge. When you learn about the 'what and why' of performance, beginning with capability measurement, you discover how much sub-optimisation or waste there is in the system. In the public sector it is significant. Paradoxically the targets and specifications regime hides the waste and adds to it both in terms of being a cause of waste, and in the creation of a costly, irrelevant and misleading bureaucracy.

Because of the problems of measures being invalid and unreliable, disputes about comparisons arise between organisations and their inspectors. Every league table that is published consumes time, increases stress, wastes energy and resources, and, most importantly, does nothing to further our understanding about how to improve. We should not treat these problems as signs that we need to improve the way we set and compare targets; we should see these arguments as symptoms of the problem – targets are of no value in understanding and improving performance.

The future of the specifications and inspection regime

My recommendation is that all targets are removed and that recent legislation associated with the current performance management regime is suspended. We should have only one requirement of public sector organisations: that they establish measures that, in their view, help them understand and improve performance. If and when they are inspected they would be required to demonstrate how they have satisfied the requirement and to what effect.

The principal advantage of this approach is that it places the locus of control where it needs to be – with those who need to change. Therefore it will be more likely to foster innovation than the current regime, which fosters compliance. It will also remove the need to comply with requirements that undermine improvement.

The savings associated with the dismantling of the specification and inspection bureaucracies will be immense. A smaller budget could be allocated to management education, guidance and support, for it remains the case that public sector management is poor. The current regime only exacerbates this problem, setting targets cannot and does not magically educate managers about method.

Finally this change will provide a framework of legitimacy for much of the current improvement work that goes unrecognised in the current regime. I am astonished at the frequency with which I come across good work being done that is not recognised because it does not fit within the inspectors' scope. There are people in the public sector who manage to improve things in spite of the system, they need a system that encourages rather than obviates improvement. That is the responsibility of Government.