

Chapter 8: Deliverology: the science of delivery, or dogmatic delusion?

In September 2006, one of the architects of the present public-service regime wrote a newspaper article in defence of the government's approach to reform¹. Sir Michael Barber, first head of the No 10 Delivery Unit, opened the piece by offering what he described as 'the global context' of public-sector reform to provide an answer. What 'global context' meant puzzled me – how can a global context help us understand how to improve public services? I was interested to find out, so I read on.

Governments, said Barber, face a productivity challenge; people want better services but don't want to pay higher taxes. To meet the challenge, he went on, three management models have emerged: command and control, quasi-markets, and devolution and transparency.

Command and control, he suggested, 'is often essential for a service which needs to improve from awful to adequate'. In support of his argument, Barber cited literacy in schools and waiting times for healthcare. He thinks of command and control as mandating attention to an issue. Whether learning in the classroom and healthcare have improved is open to question. Whether command and control is ever essential and can move an organisation from awful to adequate are also open to question. I shall return to these issues.

While claiming that moving services from awful to adequate has been a big achievement from the government's point of view, Barber accepted that the public wants 'great' service. Great service, however, could not be mandated; it had to be 'unleashed'. The second model is, therefore, a model for such 'unleashing': quasi-markets. Health and education reforms were examples: devolution of responsibility to schools, GPs, hospitals, and more choice for the consumer – parent or patient. To foster choice, the idea was to add alternative providers of these services, for example private or voluntary-sector suppliers of healthcare treatments and city academies.

The aim was to recognise that public services while different from businesses in being universal and equitable, remained similar in management terms. I was not sure he and I would mean the same by these words. Barber offered his reader an example, presumably to clarify: 'Chip-and-pin spread across the retail sector in not much more than a year.' And he asked: 'How many successful innovations have spread across public services at that kind of speed?

Any the wiser? I wasn't. What relevance has the adoption of a new security system for credit and debit cards for the improvement of public services? The only connection I can think of is the way in which new IT-dependent public services are opening services up to fraud – all the fraudster needs is knowledge of what will trigger payments as they are presenting data to a machine with rules, not a person with a brain. But I doubt Barber was making that connection.

The third model involved the 'combination of devolution and transparency'. It applied, according to Barber, in those circumstances where neither command and control nor quasi-markets were appropriate. Government contracted with or delegated responsibility to service providers, and then published the results. This would apply 'competitive pressure' and thus boost performance.

¹ 'Reform of our public services is a test for managers' Financial Times Saturday 30/9/06

The three models were apparently to be found in use around the world, making sense to an approving Barber where there was wide performance variation within a service: 'a struggling hospital with a large deficit needs command and control whereas a successful, well-led foundation hospital is best left to the disciplines of a quasi-market.'

'If this is the right approach, where is the controversy?' Notwithstanding rising public expectations, six years into Mr Blair's leadership most of the results were moving in the right direction. 'But', Barber said, 'there is more to the current challenges than that.' The critical transition from command and control required 'sophisticated strategic leadership'. What can this mean? Barber tells us that a common error is to believe that moving from command and control to a quasi-market requires 'letting go'. In fact, he said, quoting the well-known concept of David Osborne and Ted Gaebler², it requires government to 'steer rather than row'. (Osborne and Gaebler described central government's role as guiding the general direction of policy, through spending agreements and service targets, while devolving delivery and governance)

For Barber, this means a change in the role of 'officials'; but he does not elucidate who or how. Except, perhaps, for the role of the centre: the pressure from the public on government, he argues, places government in a dilemma which is best resolved by leaders putting in place excellent risk and performance management systems. And this, Barber explains, is the reason political leaders as far apart as Canada and Australia have created the equivalents of the Downing Street 'delivery unit'.

This then is what Barber offers in the way of the promised 'global context': because other countries are copying the UK, it must be the right thing to do. To close his piece, Barber informs us that public service professionals need to have the mindset and capability, not just to lead radical change but to manage transformed services. What does this mean? While many head teachers and health service managers apparently thrive with their new responsibilities, others struggle. 'It is', he says, 'all in the execution'. But Barber told us nothing about method.

So I turned to his book, *Instruction to Deliver*, to find out more. Early on he quotes the Prime Minister's frustration:

I bear the scars on my back after two years in government. People in the public sector are more rooted in the concept 'if it has always been done this way, it must always be done this way' than any group of people I have come across.³

Perhaps it was this frustration that led Blair to set up the PMDU to drive public-sector reform. It was based in his office, reporting directly to him. He gave its leadership to Barber, then plain 'Mr', who had previously been responsible for delivering the government's literacy strategy⁴. This had involved giving teachers precise instructions on how to teach what was called the literacy hour in their classrooms. Barber had delivered in education; perhaps his model could deliver for the public sector.

² Osborne, D and Gaebler, T. 1992 'Reinventing Government' Reading, Massachusetts: Addison-Wesley. See especially Chapter : 'Catalytic Government: Steering Rather Than Rowing'

³ Barber, M 2007 *Instruction to Deliver* Politico's Publishing: London P 46

⁴ See <http://www.leeds.ac.uk/educol/documents/000000153.htm>

But had Barber delivered in education? Had literacy actually improved, or had children been taught to pass tests? If it had worked, had the practice of devoting specific times in the week to teach literacy been the reason for success, or was it the method (phonics) specified? It is hard to know whether the literacy hour worked and, if so, why. Certainly we have seen plenty of evidence that ought to lead us to question whether the literacy strategy has had any real impact on children's learning.

Last year Warwick Mansell published a damning account of targets in education⁵.

[He] sets out in comprehensive and sometimes shocking detail, the pressure on teachers to deliver the improving test statistics by which the outside world judges them is proving counter-productive. Schools have been turning increasingly into exam factories... Intellectual curiosity is stifled. And young people's deeper cultural, moral sporting social and spiritual faculties are marginalised by a system in which all must come second to delivering improving test and exam numbers.

Others have also cast doubts on the achievement. For example, Durham University's Peter Tymms challenged the statistical basis behind the perceived success of the literacy strategy⁶. Tymms concluded that the statistical procedures behind the startling results on which Barber had built his reputation for delivery were faulty. When the statistical error was corrected the results flattened out. Tymms drew parallels with the US state of Texas, where similarly spectacular results had been achieved, only for 'the Texas miracle' to be revealed as an illusion. He attributed the dramatic improvements to the teachers 'teaching to the test' and concluded that the same was happening in England.

In the early days of arguing about targets with public-sector reformers, I was struck by the frequency with which people cited the literacy strategy as evidence that targets worked. Why didn't they mention the hundreds of other targets? Surely picking out (the same) one to claim the benefit for all should ring some alarm bells?

Deliverology: by what method?

By his own admission, Barber made 'deliverology' up. He developed five simple questions to be asked of any leader responsible for change⁷:

What are you trying to do?
How are you trying to do it?
How do you know you are succeeding?
If you are not succeeding, how will you change things?
How can we [the Delivery Unit people] help you?

All plausible questions.

Barber determined there were nine key issues to be addressed in order to deliver a target⁸:

⁵ Anthony Seldon in the foreword to *Education by Numbers, the tyranny of testing*, Warwick Mansell, Politico's 2007

⁶ Tymms, P. 'Are standards rising in English primary schools?' *British Educational Research Journal*, Vol 30, No. 4, August 2004 p477-494

⁷ Barber, M 2007 'Instruction to Deliver' Portico's Publishing: London P73

⁸ Barber, M 2007 'Instruction to Deliver' Portico's Publishing: London P76-7

Accountability and leadership
Project management
Levers for change
Feedback and communications
Timetable for implementation
Risks and constraints
Inter-departmental collaboration
Resources
Benchmarking

All plausible issues.

Senior civil servants were obliged to have plans covering all of these questions and issues, and these were discussed at what Barber calls ‘stocktakes’: meetings where his own Delivery Unit personnel (rather than the civil servants responsible) presented plans and progress on chosen measures, with the prime minister in attendance. Every stocktake was minuted, actions were agreed and were to be followed up. Every public-sector service subjected to deliverology had its own ‘RAG’ status. RAG stands for ‘red amber green’, an idea first developed in the private sector. Green means on target, amber signifies a few problems, red means likely to fail. You don’t want to be red.

Deliverology had an unashamedly coercive approach. Barber tells us ministers valued coercion:

David Blunkett [as Home Secretary] wanted a challenging delivery report because he thought it would drive necessary change through the Home Office⁹

Blair wanted monthly notes. Where delivery-unit people had data they used it, where not they ‘laboured to relate interesting stories’. Barber is open about the PMDU being a ‘pressure’ on reform. His argument is that focus is sufficient – people will debate it, understand it and get better at it; as they went through the deliverology routines they would get better at delivery. Or would they get better at coping with it? Would they get better at reporting, and is reporting the same as improvement?

As the ‘science’ of deliverology evolved, its management tools expanded. The trademark of management consulting ‘science’, the four-box model, arrived. It was called the ‘map of delivery’. By using a vertical axis representing ‘boldness of reform’ and a horizontal axis ‘quality of execution’, it is possible to classify change according to four types: not bold and low quality is no change (the status quo); not bold and good quality is ‘improved outcomes’; bold reform but low quality is ‘controversy without impact’; and bold reforms of high quality are labelled ‘transformation’.

The use of such a framework can only be to label things. It can be used to kick or to praise. Such actions presuppose that the assessments are reliable. But labelling doesn’t help achieve any change.

⁹ Barber, M 2007 ‘Instruction to Deliver’ Politico’s Publishing: London P118

Delivery project managers were obliged to investigate ‘delivery chains’, the hierarchies connecting the front line and thus the user experience to the minister. ‘Delivery-chain analysis’ meant involving all players that exert influence through the chain. Data and trajectories became the order of the day. Choose a place you want to be, see where you are now; draw a line. Now you can monitor your progress. But, what did deliverology teach about method, how to get there? What are the consequences of driving delivery plans down through deliver chains?

It is no surprise that deliverology itself developed delivery problems. Barber tells us that by October 2001, the position was poor¹⁰. It was decided that a harder push was needed; individuals were to be held accountable. By the end of 2002, at the time of the third round of delivery reports, progress was no better than mixed – red remained the dominant colour on the RAG status reports. Indeed, Barber reports that in 2002 most of the graphs were moving in the wrong direction. Apparently Tony Blair was becoming concerned: ‘We can’t really afford to miss as many of these targets as seems likely’¹¹.

Barber rationalised the problem as lack of focus as reforms moved from idea to implementation. He argued that civil servants were too willing to compromise if lobby groups resisted changes. He blamed bungled implementation and bureaucrats writing regulations with no thought to what practical implementation would mean. Certainly he was right about the latter. But rather than question the regime he invented – its purpose and methods – Barber reacted like anyone else with a target that they look like failing: bear down on and/or blame others.

Deliverology’s second ‘try-harder’ approach to was the ‘priority review process’. The idea was to change the focus from improving whole services to improving features of services that were political priorities, ‘sharply focused on the key delivery issues’. Policing focused on street crime and health focused on waiting times. In health, accident and emergency (A+E) was chosen as a priority because, by Barber’s account, in mid-2002 nothing was happening. Delivery Unit people found and promulgated a solution: ‘see and treat’. I assume this means putting the right expertise at the front of A+E to enable people to be seen and treated, cutting out all delays. Barber remarks that like all revolutionary ideas it was so simple that you wondered why no one had thought of it before. Unfortunately, he didn’t stick with that thought. If he had, he he might have come round to the idea that the regime itself was an impediment to effective change.

It might have occurred to Barber that the regime fostered compliance rather than experimentation. People in health are worried about target times not because of clinical considerations but because the regime is. Is this the right thing to worry about? I remember listening to a radio broadcast about the ambulance service a few years ago. In one region, the service had plotted the frequency of demand by geography – where people were likely to call for an ambulance – and moved their ambulances to those locations. Obviously assuming this was a ‘no-brainer’, the reporter moved to another part of the country and asked the managers why they did not do the same. ‘It wouldn’t work here,’ was the reply.

I reflected on this being a common phenomenon in all hierarchical organisations – the ‘not invented here’ syndrome. Good leaders know how to get around it. Instead of asking, ‘Why

¹⁰ Barber, M 2007 ‘Instruction to Deliver’ Politico’s Publishing: London P104

¹¹ Barber, M 2007 ‘Instruction to Deliver’ Politico’s Publishing: London P135

don't you do that here?' good leaders would ask (in this case): 'What do you know about demand by geography?' The answer would, of course, be not much. The leader would then say, 'Get the data, I'll be back'. If there is scope for improvement offered by understanding demand data by geography, it will become apparent. But the regime sitting on top of A+E is asking not about the nature of demand but, 'How are you doing against the targets?' So managers will focus their ingenuity on meeting the targets rather than meeting real demand. One way to do that is to reclassify conditions that attract ambulance emergency status (so fewer calls have to be met within the eight-minute target). This is the wrong thing to do and will only lead to mistakes – more people worse off.

An ambulance driver wrote to me:

My service started a process of downgrading certain types of emergencies about three years ago, ostensibly in the interests of creating responses more appropriate to the condition of the patient. But two facts give the game away: firstly, the vast majority of what management called 'regrading' were conditions being moved from the fastest (eight-minute) response-time target into categories that allowed more time. Secondly, a number of these jobs while not considered to be life-threatening necessarily, certainly could be construed to be limb-threatening or involved leaving patients in pain unattended for unacceptable periods of time (anything up to two hours depending on how much the service was under strain). Inevitably some potentially life-threatening incidents slip through the net.

Of course, many of the jobs left in the so called 'life-threatening' eight-minute response time category turn out to be anything but life-threatening as well! All this because our focus continues to be on the government-imposed target instead of studying demand from the patient's perspective.

We have been trying to bring this to the attention of senior managers and the medical director during all this time, but with little apparent result. They continue to get away with it because the public are unaware that tardy responses in these cases are not necessarily the exception and are actually designed in to a system which is basically flawed.

The gaming of classifications to meet target times is, as the correspondent suggests, simply the wrong focus. The futility of focusing on response times instead of achievement of purpose has been the subject of published research¹², but as the thinking behind the evident challenge offends the thinking behind the regime it is ignored.

In the latter days of Barber's reign, the deliverology regime shifted the emphasis from top-down command and control to what was called 'sustainable improvement, driven by the pressure of customers'. Perhaps they realised that top-down change was not sustainable, even if they could not bring themselves to accept that the whole approach was failing. Tony Blair talked of this change of emphasis as moving from 'flogging the system' to structural reform. Ministers and civil servants start describing this as 'letting go'.

¹² 'Treating the clock and not the patient: ambulance response times and risk' L. Price, Qual. Saf. Health Care 2006; 15; 127-130.

This was in 2004, more than three years into deliverology. For Blair this shift was a new vision of reform, involving higher standards of performance through greater customer responsiveness. The tailoring and personalisation of services, services built around customers, not producers, as usual sounded plausible. But paradoxically, it was the regime that had created the offending producer interest in the shape of the specifications industry which grew around deliverology and dictated the measures and organisational designs: the very programme that is designed to ‘deliver’ delivers the programme, not reform. This is why, in 2004, while the regime’s numbers were going up public satisfaction with services was heading in the other direction¹³.

The shift of emphasis meant greater political focus on ‘choice’ and competition. It also meant involving citizens in the design and delivery of public services. As I shall describe later, this shift in emphasis to citizen-engagement has led to some very silly consequences.

Barber describes deliverology as ‘world-class tools and processes’. I think of it as Mickey Mouse command-and-control. That is being generous to Barber and unfair to the mouse. Deliverology’s method amounts to determining change on the basis of opinion and driving activity down into systems with no knowledge of the impact on the way the system will perform. It is tampering on a massive scale.

The plausible ideas – the questions and issues posed by deliverology – were not ideas that would foster knowledge; quite the reverse, they would foster planning, project management, reporting, rationalisation and other dysfunctional behaviours. Like all command-and-control regimes, deliverology tries to control the behaviour of people, to reduce the chances of them behaving opportunistically, but, paradoxically, it will be more likely to create and amplify such behaviours.

To review Barber’s logic: firstly, what he calls the productivity challenge – people wanting better services, without paying higher taxes. This assumes, as a command-and-control thinker would, that more resources is the only way to better services. Yet the better way to improve productivity is by improving quality, as Deming taught. Ohno put it this way: if the capacity of the system is the work and its waste, the way to improve capacity is to get rid of the waste. And, as I have shown, the waste is caused by the regime. People want better services *and* don’t want to pay higher taxes.

Barber believes that targets, what he thinks of as a command-and-control approach, helps services improve from awful to adequate. It is a rationalisation, based more on the implicit assumption about the unwillingness of public-sector managers to act than the evidence provided by managing with targets. Having rationalised the state the public sector is in following the imposition of targets, Barber now claims that services need a different type of intervention to go from adequate to good. Why should that be the case? How can two different strategies for improvement be relevant to any starting-place?

The consequences of ‘devolution and transparency’ – delegating services to contractors with service-level agreements – are alarming. The regime has set up factories to provide – among other things – healthcare, legal advice, consumer advice and local-authority services. In many cases, the contracts assumed (and paid for) a level of demand that was not forthcoming – an incredible waste of public funds. Worse, as the regime measures these factories on

¹³ Barber, M 2007 ‘Instruction to Deliver’ Politico’s Publishing: London P236

transaction costs, they are deluded into assuming lower transaction costs means better performance. I shall return to the public-sector factory problem in Chapter 11.

Barber reported the Prime Minister's worries about not hitting targets. One key target was the number of asylum seekers. Between April and June 2004, the number of asylum seekers fell by 13 per cent, and the number of failed asylum seekers being deported also fell. Did they fall because the process had been improved, or were more people allowed to enter so that the government could meet its target? The shadow Home Secretary, David Davis, accused the government of handing out work permits 'like confetti' to meet the targets on asylum¹⁴. I remember news reports at the time of insiders describing how they had been instructed to change the rules (to admit people); if true it would come as no surprise to a systems thinker. The targets bureaucracy takes over the management of the work; the focus becomes meeting the targets rather than improving the way the work works.

Like many who write about organisations, Barber cites Norman Dixon, a psychologist who wrote *On the psychology of military incompetence*¹⁵. Dixon's thesis is that military organisations knock the attributes of leadership out of people as they progress through the hierarchy, so that by the time they reach the top they are no longer fitted to leadership. Barber uses Dixon to draw a parallel with the public sector, alleging that public servants betray a 'fundamental conservatism', a 'tendency to reject or ignore information which is unpalatable', and 'an obstinate persistence in a given task despite strong contradictory evidence'.

It is arguable that it is Barber himself who has been brought to 'incompetence' by ploughing onwards with nothing more than a set of plausible ideas, ignoring evidence of their lack of success and showing his own persistence in the face of contradictory evidence. One distinguishing feature of senior public-sector managers is their intelligence. Is it conceivable that resistance to deliverology might have been a reasonable reaction to stupid ideas? Any such resistance would have been amplified by being obliged to sit through a third party's assessment of your organisation delivered in front of the Prime Minister.

Barber claims that an academic article by Steven Kelman¹⁶ debunks the urban myths about targets. Careful reading shows nothing of the sort. Kelman sets out to relate the PMDU's activity to concepts in organisational change theory. His empirical work consists of interviews with people involved. The only conclusions that are evidence-based are his findings that people working in the PMDU were more likely to be orientated to imposing practices, were less convinced of the front-line people's commitment to targets, and were more sceptical about lower-level willingness voluntarily to embrace performance improvement using targets. In other words, PMDU staff were more of the view change had to be driven by the centre.

Barber's 'deliverology' falls short on method. It is only method that can answer the question: how can services be improved? Barber's *de facto* method is to create a bureaucracy for measuring and reporting that then deludes people into assuming improvements are real; his strategies for 'unleashing' only unleashed diseased and dysfunctional bureaucracies. One

¹⁴ <http://www.guardian.co.uk/immigration/story/0,,1422806,00.html>

¹⁵ 'On the Psychology of Military Incompetence', Norman Dixon, Futura Books, 1976.

¹⁶ Kelman, Steven (2006) 'Improving service delivery performance in the United Kingdom: Organisation theory perspectives on central intervention strategies', *Journal of Comparative Policy Analysis: Research and Practice*, 8:4, 393-419.

service where we have seen the direct effect of deliverology is the police service. Deliverology caused the establishment of bureaucracies for crime recording and reporting; as we shall see, this is not the same as policing.

Copyright Triarchy Press