

Keeping **Up** the standards?



The ground may be shifting on the use of targets and rankings to improve performance. **Ruth Dixon, Christopher Hood** and **Deborah Wilson** look at the past and future of the standards agenda

Targets, benchmarks and league tables have a long history in education and in other public services, and concern about the effects of such performance indicators (PIs) is also not new. As far back as 1874, the Reverend John Menet wrote a pamphlet savagely criticising newly introduced School Standards (a form of payment by results whereby government grants to elementary schools were based on the number of children passing examinations ('Standards') of different levels (Aldrich 2000)). He asserted that the whole system was "radically vicious" on eight counts (see figure 1; taken from Menet (1874)) such as the neglect of high and low achievers and the encouragement

of "a mechanical routine in teaching". Menet argued that Standards tied the hands not only of school managers but also of inspectors, since the Standards became almost the only basis of judging a school, thus removing the autonomy of inspectors to encourage teachers to go beyond the basics and "aim for the highest possible standard of instruction". The reverend gentleman was not exactly a dispassionate observer: as the first principal of Hockerill Teacher Training College (founded 1852) in Bishop's Stortford, he might have been expected to wish to assert the independence of teachers from a system of national standards. But he pointed to significant problems which resonate today.

Was he right? Should we discard the modern-day equivalent of nineteenth century School Standards such as SATs, GCSE targets, contextual value added scores and school league tables? All of these have their critics and consume time and resources within schools, local authorities and the school inspectorate. Should we even go a step further than Menet suggested and do away with school inspection altogether? After all, Ofsted costs the equivalent of about 5,000 teachers – couldn't it be scrapped to release resources to go into classroom teaching by a future cost-cutting government?

The resources expended on meeting standards or targets can only be justified if it is true that (as has been claimed) performance management by numbers can transform educational quality. So what is the evidence that 'managing by numbers' improves public services? And if PIs are to be retained for the coming age of public service austerity, how could they be used better in the future?

Do performance indicators improve public services? Policy differences across the countries of the UK post-devolution, with England relying more

CONTENTS.		PAGE
INTRODUCTION. Indications that alterations of the Code are at hand ..		5
I. Examination by Standards. Reasons why the system is radically vicious		6
1. Standards must apply to all Schools, scarcely any two being alike		7
2. Standards, being of necessity a <i>minimum</i> , become a <i>maximum</i> ..		8
3. Standards discourage the progress alike of the clever and dull Children		8
4. Standards encourage a mechanical routine in teaching		9
5. Standards tend to turn Schools into commercial ventures; their effect in this aspect on Managers, Teachers, and Children ..		16
6. Standards interfere with the proper organization of a School by classes		18
7. Standards known only in Elementary Schools, and opposed to the practice of all other Schools and to common sense ..		18
8. Standards tend to the substitution of Examination for Inspection		19
II. Outline of remedies proposed :		
1. Abolition of Standards		21
2. Full Inspection with Examination		21
3. Payment by Class		21
III. The Pillars upon which a system of Education must rest, viz. :		
1. A due supply of Trained Teachers, to be secured by the Pupil Teacher system and the Training Colleges		22
2. Inspection		22
Experience to be gained from the Minutes of 1846		22
IV. Inspection. Consideration of some question connected with its efficient working		23
Conclusion		26



The Reverend John Menet. Reproduced by kind permission of Bishop's Stortford Museum at Rhodes.

Figure 1. Contents page from Menet (1874) *The Standards of the New Code*. Reproduced by kind permission of the University of Bristol Special Collections and the JISC Digitisation Programme which made possible the Nineteenth Century Pamphlets Project of Research Libraries UK.

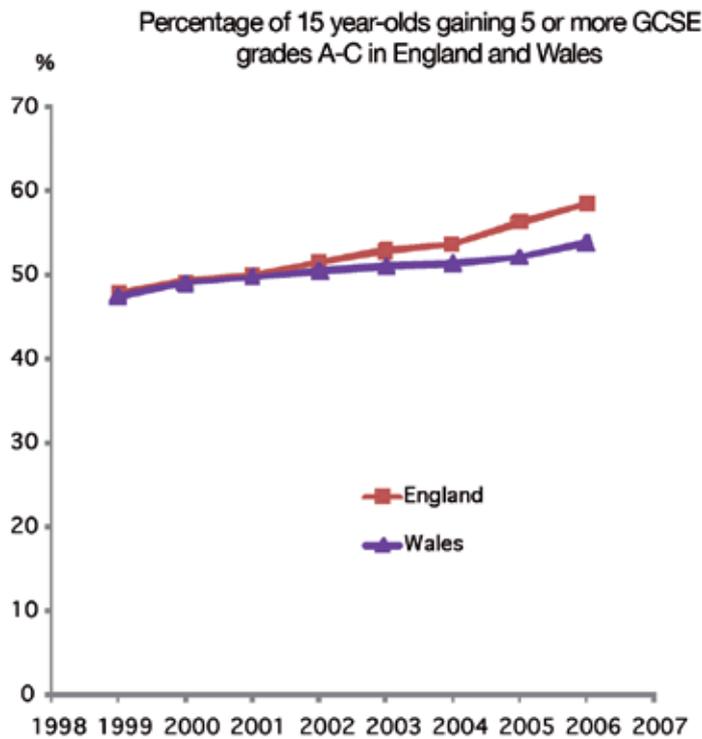


Figure 2a. GCSE results from England and Wales

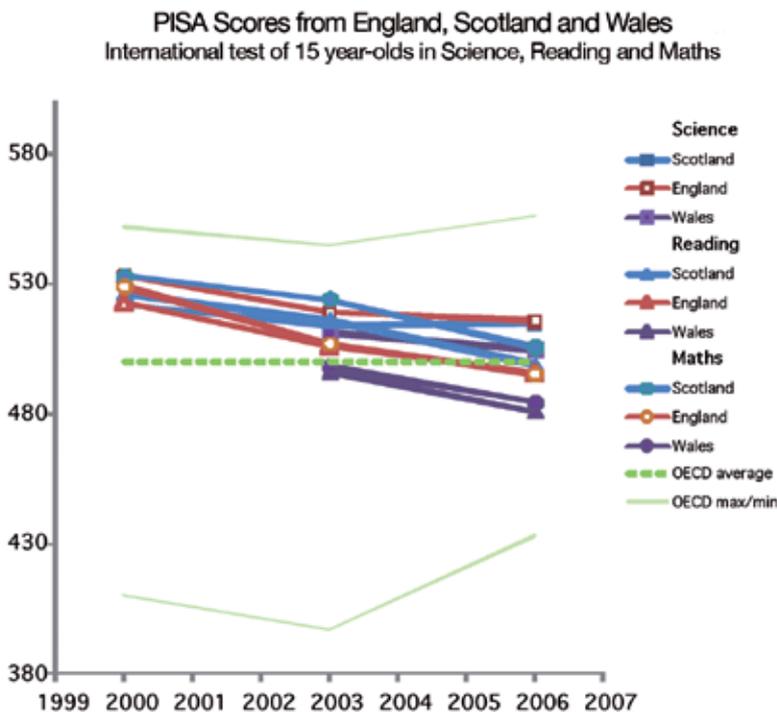


Figure 2b. PISA scores from England, Scotland and Wales

heavily on PIs than its neighbours, give us one opportunity to investigate their effects. In some policy domains, high-stakes PIs do appear to have improved public services. For example, there is evidence that waiting times for elective surgery in hospital fell faster in England than in Scotland in the early 2000s, when the English PIs were linked to a tough regime of sanctions for target failures that was not present in Scotland at that time (Propper *et al.* 2008). In education similar patterns appear. Wales scrapped school league tables in 2001 while England retained them, and figure 2a shows that there was a more rapid increase in the percentage of pupils gaining five or more GCSEs at grades A*-C in England than in Wales between 1999 and 2006.

Different pattern

While such comparisons of UK countries would suggest that high-stakes PIs can help to drive public service improvement both in health and education, international comparisons show a different pattern. Every three years the OECD tests a large sample of 15-year-olds from over 40 countries in its Programme for International Student Assessment (PISA). On those tests, all the UK countries have shown a decline in performance, both in absolute scores and relative to other OECD countries, and England does not stand out from Scotland, Wales or Northern Ireland on this measure (figure 2b illustrates this for England, Scotland and Wales).

How do providers respond to PIs? Clearly, performance management regimes are designed to influence the behaviour of service providers in order to produce 'desirable' outcomes. But they can also influence providers' behaviour in less beneficial ways. The main forms of strategic or gaming behaviour that can result from target systems are ratchet effects, threshold effects and output distortions (Hood 2006 and 2007; Bevan and Hood 2006). Ratchet effects arise in systems where the target is set as an incremental

advance to the current output, which creates the incentive to suppress that current output in order to reduce future targets. Threshold effects occur due to the incentive to just hit the target and do no more; and output distortions describe the incentive to focus on hitting the target to the detriment of effort directed at real improvements in performance.

Such observations are far from new: Reverend Menet described many of these effects in the school system in 1874 and examples still abound across the public services. A clear example of a threshold effect is the finding by Burgess *et al.* (2005) that outcomes for low-achieving children were adversely affected by the introduction of the five A*-C GCSE target and the resultant focus of resources on borderline pupils. This finding was corroborated by one of us (Wilson *et al.* 2006) on the basis of in-depth interviews with 21 secondary school headteachers, which revealed that 11 had at some point targeted resources at pupils on the C/D borderline, to the potential disadvantage of children above or below that level, although six stated that they deliberately did not follow that strategy.

Next we must consider: are performance indicators reliable

“How much faith should a parent choosing a school for their 11-year-old child today have in the most recent CVA data, which is based on the GCSE performance of pupils starting school up to seven years before?”

measures? And are rankings meaningful? The extent to which PIs can lead to better public services also depends on how well the indicators capture the aspects of performance that policymakers want public service providers to focus on. If PIs are unstable or unreliable, the public cannot be sure that performance is being meaningfully measured and public service providers have no firm basis for assessing their performance. Indeed, Peter Tymms (2004) has forcefully argued that as both the National Curriculum for England and Wales and the form of SATs have evolved over time since their introduction in 1996, it is very difficult or even impossible to assess what changes in test scores mean.

Uncertainties

Even if we step back from that problem, for PIs to be used in meaningful league tables, they need to discriminate clearly between different units (schools) and

to complement one another. Many rankings ignore the uncertainties associated with measurement, even though Deborah Wilson and Anete Piebalga (2008) have shown that over half of all English secondary schools are not significantly different from the national average when ranked on their contextual value added (CVA) scores. And how much faith should a parent choosing a school for their 11-year-old child today have in the most recent CVA data, which is based on the GCSE performance of pupils starting school up to seven years before? Can they predict the future performance of a school from this information?

George Leckie and Harvey Goldstein (2009) compared CVA scores for the same schools from 2002 and 2007 in order to estimate the uncertainty associated with projecting performance scores five years into the future (the maximum time for which comparable data were available). They found that after taking account of this uncertainty, 97 per cent of schools were not significantly different from the average. As Leckie and Goldstein say in their paper, “using current school performance as a guide to future school performance is highly misleading”. Further, the same organisation can have different positions in a ranking exercise depending on which aspect of performance is measured. Wilson and Piebalga (2008) showed the extent to which ranking English secondary schools on uncorrected (5A*-C) GCSE scores produced a very different league table to ranking them on CVA scores.

All that might suggest that there might be much to gain, especially in an age of austerity, from shifting the resources that go into PIs and



the associated testing regimes into teaching. After all, England is now the only country in the UK that retains official school league tables, and its performance on the PISA ratings is not obviously better than other parts of the UK. Even in England, only one out of the SATs survives, and the future of that has been called into question in recent months by statements from both the main political parties (which would no doubt have pleased Reverend Menet). Given that the British public apparently has little trust in official statistics (a Eurobarometer survey three years ago ranked the UK lowest of all EU countries on public trust in government statistics), it might be argued that the ranking of schools might be better done by the press or private ratings organisations rather than by state bureaucrats.

But the idea that scrapping public testing and targets would release a lot of resources which could go into hiring thousands of extra classroom teachers seems naive. After all, even if public testing were abandoned, teacher assessments would remain. Nor would scrapping PIs remove the need for qualitative school inspection

– indeed, it might increase that need. And at a time of approaching fiscal austerity, when the targets that will really matter for everyone are likely to be targets for input reduction and productivity increases, it may be all the more important to have comparable measures of non-financial performance.

Intelligence

Moreover, PIs can have a value for ‘intelligence’ – information that can be used to inform policy or practice – even and perhaps especially when they are not coupled with targets or rankings. Indeed, when PIs are used in this way, they have no predictable effects for providers and so are hard to ‘game’. For example, an absolute categorisation of schools on the basis of their CVA score enables those at both extremes of the performance distribution to be identified, which provides a starting point for a dialogue in which schools have to account for their performance. This seems a more sensible use of this performance data than the largely spurious ranking of every individual school that is the current focus.

Similarly, the development of better ways of comparing performance

data across the different countries of the UK could lead to more informed comparisons than are easily possible with the current system, where there is a high degree of incommensurability in the indicators used by different countries.

So the key challenge for the 2010s would seem to be more one of how to use PIs more intelligently in the context of better performance regimes than of doing away with them altogether. More intelligent use of PIs might fruitfully fulfil Reverend Menet’s 1874 call for “less routine, less mechanism”. They might not achieve his desire for “less complication, fewer pains and penalties”, but that is probably not achievable in a modern education system under pressure to perform and in a world where complexity cannot be avoided.

Ruth Dixon, Christopher Hood and Deborah Wilson work for the ESRC Public Services Programme (www.publicservices.ac.uk). This article is based in part on their recent paper ‘Managing by Numbers: the Way to Make Public Services Better?’, available at www.publicservices.ac.uk/wp-content/uploads/managing-by-numbers.pdf

References

- Aldrich, R (2000), ‘Educational Standards in Historical Perspective’, in Goldstein, H; Heath, A (eds), *Educational Standards. Proceedings of the British Academy* 102: 39-67
- Bevan, G; Hood, C (2006), ‘What’s measured is what matters: targets and gaming in the English public health care system’, *Public Administration* 84(3): 517-538
- Burgess, S; Propper, C; Slater, H; Wilson, D (2005), *Who wins and who loses from school accountability? The distribution of educational gain in English secondary schools*, CMPO discussion paper 05/128, University of Bristol
- Hood, C (2006), ‘Gaming in targetworld: the targets approach to managing British public services’, *Public Administration Review* 66(4): 515-521
- Hood, C (2007), ‘Public service management by numbers: why does it vary? Where has it come from? What are the gaps and the puzzles?’, *Public Money and Management* 27 (April): 95-101
- Leckie, G; Goldstein, H (2009), ‘The Limitations of Using School League Tables to Inform School Choice’, *Journal of the Royal Statistical Society series A* 172(4): 835-851
- Menet, J (1874), *The Standards of the New Code*, Bristol Selected Pamphlets, www.jstor.org/stable/60244630
- Propper, C, Sutton, M, Whitnall, C; Windmeijer, F (2008), ‘Did “Targets and Terror” Reduce Waiting Times in England for Hospital Care?’ *B. E. Journal of Economic Analysis and Policy* 8: 2, article 5, www.bepress.com/bejeap/vol8/iss2/art5
- Tymms, P (2004), ‘Are Standards Rising in English Primary Schools?’, *British Educational Research Journal* 30(4): 477-494
- Wilson, D; Piebalga, A (2008), ‘Performance Measures, Ranking and Parental Choice: An Analysis of the English School League Tables’, *International Public Management Journal* 11(3): 344–366
- Wilson, D; Croxson, B; Atkinson, A (2006), ‘“What gets measured gets done”: headteachers’ responses to the English secondary school performance management system’, *Policy Studies* 27(2): 153-171