

[Gwyn Bevan](#) and [Barbara Fasolo](#)

Models of governance of public services: empirical and behavioural analysis of 'econs' and 'humans'

Book section

Original citation:

Originally published in Bevan, Gwyn and Fasolo, Barbara (2013) *Models of governance of public services: empirical and behavioural analysis of 'econs' and 'humans'*. In: Angus, Oliver, (ed.) *Behavioural Public Policy*. Cambridge University Press, Cambridge, UK, pp. 38-62. ISBN 9781107617377

© 2014 [Cambridge University Press](#)

This version available at: <http://eprints.lse.ac.uk/59185/>

Available in LSE Research Online: September 2014

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

This document is the author's submitted version of the book section. There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

Models of governance of public services: Empirical and behavioural analysis of 'econs' and 'humans'

Gwyn Bevan and Barbara Fasolo

Published as Chapter 2 in Oliver, A. (Ed.). (2013). *Behavioural public policy*. Cambridge University Press: 38-62.

Models of governance of public services: Empirical and behavioural analysis of 'econs' and 'humans'

This chapter develops Le Grand's argument about the need to recognise that those who choose to work in public services are not wholly 'knights' (Le Grand, 2003) driven by *altruism*, but are a mix of 'knights' (altruistic) and 'knaves' (selfish). He argued that *choice and competition* between schools and hospitals through quasi markets in which money follows the pupil or patient is the primary policy instrument to combat government failure. This is because quasi markets appeal to both knightly and knavish motives. They create pressure to improve services to respond to threats to market shares, providers' incomes and hence jobs (Le Grand, 2007). The reason why we disagree with Le Grand in his proposed remedy to his diagnosis is that the choice and competition model assumes that people as users and providers of public services act, as 'econs', as described by Thaler and Sunstein (2008): i.e. behave as in conventional micro-economics. We know that 'humans' behave differently, as has been demonstrated by carefully designed psychological experiments. Thus, although, if asked, people will say they desire more choice and believe that the greater the choice the better, there is ample evidence that in health care, just as in more mundane consumer choices (Iyengar and Lepper 2000), users do not behave like econs and do not use information on provider performance to switch from those that are poor to those that are good. Instead of the choice and competition model, we see scope for developing the more psychologically plausible model of governance based on users and providers as 'humans', where users do not use public information on provider performance to switch between providers, but providers respond to threats to their *reputation* via public reporting of rankings of performance (Hibbard et al. 2003).

In this first section we present the four alternative models of governance: altruism, hierarchy and targets, reputation and choice, and competition.

- *Altruism* assumes providers are 'humans' and internally motivated to perform well. To do better, they need more resources or information. This model does not require external incentives, has low monitoring costs and is popular with professionals (Le Grand, 2007). In the National Health Services (NHSs) of the UK this was the traditional model and associated with a system where failure was rewarded and success ignored (Bevan, 2010).
- *Hierarchy and targets* assumes providers are 'econs' and respond to rewards for success and sanctions for failures. This model imposes external incentives by strong performance management, has monitoring costs and is unpopular with professionals (Le Grand, 2007, Bevan and Hood, 2006).
- *Reputation* assumes providers are 'humans' and respond to threats to their reputation. A reputation model is a system of performance measurement that satisfies criteria specified by Hibbard et al. (2003): a ranking system, published and widely disseminated, easily understood by the public (so that they can see which providers are performing well and poorly), and followed up by future reports (that show whether performance has improved or not). This model 'names and shames' providers that perform poorly, has monitoring costs and is unpopular with professionals.

- *Choice and competition* assumes users and providers are 'econs' and that users choose better performing providers, whereas providers respond to the consequences of these choices on market shares. This model creates external incentives by a quasi market system in which there is choice of providers and money follows the pupil or patient. It is difficult to design effective quasi markets as they require good information, supply-side flexibility and freedom to manage. Quasi markets have high transaction costs, but are popular with governments, because pressure on poor performance comes from an 'invisible hand' (Le Grand, 2007). They promise to have more potential to respond to users' needs than centrally-driven systems of hierarchy and targets and reputation.

The next section of our chapter reports empirical evidence of the impacts of these different models using a rich set of contrasting examples described below. These examples show that the applicability of each model importantly depends on the accountability structure embedded in the organisation, and, that often, in practice, the models other than altruism may be applied in various combinations.

- *US hospitals* are not generally directly accountable to government and have to generate their incomes from market shares in a system in which there is enough excess capacity for hospitals to compete. These hospitals may be governed by altruism, reputation or choice and competition.
- *The four NHSs of the UK* are accountable to Ministers in each country, and hence all four models of governance are applicable. From 1997, all governments abandoned the model of competition (the 'internal market') in favour of altruism. From 2000, the government in England abandoned altruism and introduced governance by hierarchy and targets and reputation (through 'star rating'): hierarchy and targets, as chief executives of 'zero-rated' organisations were at risk of being sacked; and reputation, as 'star ratings' were annually published in national and local media. From 2006, the government in England added policies based on choice and competition. The governments of the devolved countries (Scotland, Wales and Northern Ireland) continued with policies based on altruism.
- *Secondary schools in the UK* are either run by elected local councils or are independent. British schools are not under direct central government control: the Secretary of State for Education is not empowered to sack a headmaster of a 'failing' school. This means that Ministers cannot apply hierarchy and targets and the alternatives are hence altruism, reputation, or choice and competition. The government in England has continued to govern through choice and competition (with quasi markets) and reputation (by publishing league tables for examination results for 16-year olds) since 1994. The government in Wales uses quasi markets only, having stopped publishing league tables in 2001. The government in Scotland relies on altruism: there has never been a quasi market and publishing league tables stopped in 2002.

We find that: reputation and hierarchy and targets were effective drivers of improvement for hospitals and ambulance services in the UK; reputation was effective for hospitals in the US and schools in the UK; and that neither altruism nor choice and competition were as effective as the other models. This raises the familiar challenge to economics: If the reputation model works in practice, can we show how it works in theory? The third section draws on behavioural economics to explain the psychological mechanisms behind each model. We conclude by considering the paradox:

why do governments in the UK prefer to use models based on altruism or choice and competition when practice and theory suggests that these are relatively ineffective?

Effectiveness of the four models of governance: 3 case studies

1. US Hospitals

In their systematic review of evaluations of public reports of hospital performance, Fung et al (2008) found that the impacts of publishing performance varied, and concluded that the choice and competition model was sometimes effective. Hibbard (2008) questioned this conclusion as she argued that the effect of report cards could be due, not to choice and competition, but rather to reputation citing evidence from a controlled experiment (Hibbard et al. 2003, 2005). This experiment was designed to examine the impacts of reporting on quality of care across three sets of hospitals: public-report, where the report was disseminated widely to the public; private-report, where the report was supplied to managers only; and no-report, where no information was made available publicly or privately. They found that the public-report set made significantly greater efforts to improve quality than the other two sets, and that threats to reputation and not market shares¹ were the key driver of performance. Private reporting, which relied on altruism, had proved to be a weak driver of change. This finding is supported by evidence from the most studied system of public reporting (Marshall et al. 2000; Fung et al 2008), namely the Cardiac Surgery Reporting System (CSRS) of the New York State Department of Health. CSRS satisfies Hibbard's criteria, with one qualification: hospital performance was not ranked but 'outliers' (with mortality rates statistically significantly higher or lower than the mean) were identified. One of the paradoxes of this system, which questions the efficacy of the model of choice and competition, was that neither users nor providers acted as 'econs'. Users continued to go to those hospitals that CSRS showed had significantly high risk-adjusted surgical mortality rates. Although poor performance had no effect on providers' market shares, they did respond to threats to their reputations: 'When a hospital is publicized as having the worst mortality in the state, not only do physicians and hospital administrators pay attention, but there also is a greater likelihood that the resources necessary to correct the problem will be forthcoming' (Chassin, 1996, p. 88). The outcome was that New York State had 'the most rapid rate of decline (of risk-adjusted mortality rates) of any state with below-average mortality' (Chassin, 2002)².

2. The four NHSs of the UK

The evidence from the US suggests that the reputation model has greater impacts than those of altruism and choice and competition. But this evidence is limited: it is from one exercise across a small number of hospitals in Wisconsin and cardiac surgery in New York State. The NHSs in the four

¹Later analysis showed that that these managers were correct: 'There were no significant changes in market share among the hospitals in the public report from the pre to the post period ... no shifts away from low-rated hospitals and no shifts toward higher-rated hospitals in overall discharges or in obstetric or cardiac care cases during any of the examined post-report time periods' (Hibbard et al., 2005).

² There are issues of gaming in response to CSRS and this claim is disputed (Dranove et al. 2003).

UK countries following devolution offer evidence of the impacts of the four different models on a massive scale. Research into the choice and competition model has failed to find evidence of powerful direct effects and econometric studies have estimated its impacts to be limited³. In contrast, as we now report, the effects of the policies based on hierarchy and targets and reputation models were direct and the impacts dramatic. We start from presenting evidence from ‘star ratings’, which applied in England from 2000 to 2005. The revolutionary idea of the annual ‘star rating’ of NHS providers was that success would be rewarded and failure would result in sanctions. This was emphasised with the threat of the sack for failure, to make clear to those who worked in the English NHS the sharp break from policies of the past. Once this was well understood, and the power of the reputation model became clear, there appears to have been less emphasis on sacking chief executives. So although the ‘star rating’ system combined the models of hierarchy and targets with reputation, over time, the model shifted from hierarchy and targets to reputation.

This is a good ‘natural experiment’ to compare the impacts of different models between the NHS in England and the NHSs in the devolved countries because:

- Each country had similar systems of health care, received a sustained massive influx of ‘growth money’, and pursued similar targets for reducing hospital waiting times and reducing ambulance response times to calls for life-threatening emergencies;
- England sought improvement through hierarchy, targets and reputation;; and
- The devolved countries sought improvement through altruism.

We consider the period from 1997 to 2005, which offers two different kinds of comparisons:

- A ‘before and after’ comparison for England of altruism (1997-2000) vs. ‘star ratings’ for acute hospital trusts (2000 to 2005), and ambulance trusts (2002 to 2005) (akin to a ‘within-subject’ study in experimental psychology);
- A comparison over the period of ‘star rating’ between hierarchy and targets and reputation in England, vs. altruism in the devolved countries (akin to a ‘between-subject’ study).

Figure 1 shows the transformation in performance of the English NHS in terms of the reduction of numbers with long waits for hospital admission following the introduction of ‘star ratings’ from 2000. The targets in star ratings became more demanding over time and were that no one would be waiting more than 12 months by April 2003 and more than 9 months by April 2004. These targets were achieved. Figure 2 shows comparative performance across the four countries of the percentages of those waiting more than more than 12 months for inpatient elective hospital admission between 2000 and 2003: only in England did this percentage fall from what it had been in 2000 when each country’s NHS began to experience substantial real growth in resources. Figure 3 gives numbers per thousand waiting more than 6 months for inpatient admission and shows how the improvement in performance for reducing this number for Wales (and Northern Ireland) lagged

³ We are aware of recent econometric studies (Cooper et al, 2010; Gaynor et al, 2010) that have identified improvements in quality, with a common measure being mortality rates from Acute Myocardial Infarction (AMI) in areas where competition has increased. We are puzzled by these findings, given that other studies showed that patients rarely exercised choice between hospitals; and that even if they had done so, this would be for elective care, whereas AMI mortality is a measure of the quality of emergency care.

that for England⁴. Figure 4 gives numbers per thousand waiting more than three months to be referred to a specialist, and shows that whereas in England this ratio was also reduced, in Wales (and Northern Ireland) these ratios increased. Hospital performance influenced the targets each country set their NHSs.

The comparison between England and Wales is best in terms of a 'natural experiment' as these countries were most similar in policies and organisation prior to devolution and although spend per capita in Wales exceeded that of England, the degree of excess was less than that for Northern Ireland and Scotland. Besley et al (2009) undertook a rigorous econometric analysis comparing data on waits for inpatient admission in England and Wales, and found that, given their starting positions in 2000 the performance in Wales was worse than in England. The stark differences between these targets as set for England and Wales was highlighted by the Auditor General for Wales (2005) and are summarised in Box 1. This shows that by December 2005, at the end of the period of star ratings in England, to achieve the targets for specialist referral and hospital admission, hospitals in England would have to complete both within nine months, but within Wales would have been allowed *three years*.

Hospitals are an exemplar of the complexities of multi-tasking, as identified in the landmark paper by Holmstrom and Milgrom (1991); hence evidence on performance on waiting times is problematic as these are inadequate measures of their performance⁵. The next example from the NHS is not vulnerable to that objection: for ambulance services their overriding priority ought to be to respond as quickly as possible to what appear to be life-threatening emergency calls. Hence if these organisations were driven by altruism they would aim to do this without any need for external incentives.

The governments in England, Wales and Scotland set a target for their ambulance services of meeting 75% of category A calls within 8-minutes: in England and Wales in 2001, and in Scotland by 2007-08. Figure 5 shows the distribution of performance for ambulance services in England before and after the introduction of their 'star ratings' in 2002 (one year after they had been introduced for acute hospital trusts) in terms of the percentage of category A calls met within 8 minutes. In 2001, only one service met that target, with many services responding to less than 50% of category A calls within 8 minutes. After the introduction of 'star ratings' all services improved and in 2003, each service either achieved or was close to achieving that target. Figure 6 compares national performance in terms of the percentage of category A calls met within 8 minutes across England, Wales and Scotland. This shows the striking difference between countries for those waiting for an ambulance to respond to what may be a life-threatening emergency: in England about one in four can expect to wait longer than eight minutes, but in Wales and Scotland this proportion is almost

⁴ After 2003, the definitions of those on waiting lists changed in Scotland to exclude from their statistics up to a third of those who were waiting for admissions and the data for Scotland are not comparable with other countries. But using comparable data for the time patients waited after being discharged, Propper et al (2008, 2010) undertook a rigorous econometric analysis of Scotland and England and showed that using these data performance in Scotland was worse than in England.

⁵ This objection is qualified to some extent by the way in which 'star ratings' were based on a set of other indicators in a 'balanced scorecard' (see Bevan and Hood?, 2006); and that, for their first three years, 'star ratings' of acute hospitals also included assessments of their implementation of clinical governance (Bevan and Cornwell, 2006). Furthermore, the public's principal and consistent complaint about the NHS in the 1990s and early 2000s was its long waiting times. Moreover, as Connolly et al (2010) point out there is no evidence of better performance by hospitals in the devolved countries in comparison with England on other dimensions of quality.

one in two. Bevan and Hamblin (2009) contrast responses by the governments in England and Wales to failure by their ambulance services to meet the category A 8-minute target. In England, that failure resulted in public censure and sackings of chief executives of ambulance services. In Wales, the failure resulted in the government setting successively less demanding ‘milestone’ targets: from April 2004, the target was reduced to 65% (the threshold for a service in England to have been zero rated); from April 2005 to 60%.

3. Secondary schools in the UK

The models of school governance in England in many ways foreshadowed the models applied later into the NHS (see box 2). These included the introduction of quasi markets in which ‘money followed the pupil (or patient)’, introduction of national standards, the creation of new inspectorates to inspect all organisations over a four-year period, and the publication of comparative performance on a national basis. As these policies were introduced prior to devolution, they also broadly applied to Wales⁶. Scotland has always had a distinctively different educational system with markedly different policies. Teelken (2000) reviews differences between England and Scotland highlighting the comparative lack of diversity of schools in Scotland and the absence of a national curriculum, and a weak quasi market. Box 3 summarises key differences for schools’ policies across the three countries. For our purposes, the key differences are in the publication of school league tables of examination performance for secondary schools at age 16 (the normal school leaving age):

- In England, the government has published this information every year from 1994.
- In Wales, the government published this information every year from 1994 until 2001 but not thereafter.
- In Scotland, the government published this information every year from 1998 to 2002 only.

The government’s league tables for England, published in 2009, give school performance by each local authority over four years (from 2006 to 2009) for the percentage of pupils at the end of Key Stage 4 (i.e. taking GCSE between ages 14 and 16) achieving five or more grades above C; and the averages for the local authority, and England. This is designed to offer information by local authority and hence for choice and competition. The *Daily Telegraph* ranks school performance by county⁷. Such school league tables satisfy Hibbard’s four criteria⁸ and do indeed have reputational effects.

Figure 7 shows how throughout this period, the ratios of pupils to teachers have been similar in England and Wales (in 2004, an average secondary school teacher in each country would have had about 17 pupils), although this ratio has recently fallen in England but remained constant in Wales; and that Scotland has throughout had a much more generous ratio (in 2004, an average secondary school teacher would have had about 13 pupils). Figure 8 shows examination performance in terms

⁶ Although Wales, unlike England, did not publish league tables of test results for primary schools.

⁷ <http://www.telegraph.co.uk/education/leaguetables/6974678/GCSE-league-tables-Key-stage-4.html>

⁸ These tables are a ranking system (although this is not done by the government), the information is published and widely disseminated, it is easily understood by the public who can see which schools are performing well and poorly, and is published annually.

of five good grades at age 16 improving in all three countries from 1994 to 2002. In 1994, Scotland had the best performance (48%), England next (43%) and Wales the worst (39%). In 2002, Scotland had improved and still had the best performance (60%), England and Wales were similar (50%). After 2002, performance in Scotland slightly deteriorated and although performance improved in both England and Wales the gap between them re-emerged. In 2008, England had the best performance (64%) and Wales had similar performance to Scotland (58%).

Burgess et al (2010) have undertaken a rigorous econometric analysis of the natural experiment between Wales and England and found 'systematic, significant and robust evidence that abolishing school league tables markedly reduced school effectiveness in Wales'. They estimated the relative impacts of the policy in Wales to have been: a fall of nearly 2 GCSE grades per student per year, and for Welsh schools to achieve English levels of performance they would need Welsh class sizes to be about 30% smaller than in England. This effect could be due to reputation, or to choice and competition; however, these same authors also point out that the different impact of policies does '*not vary significantly by the level of local competition*', and because there is generally a relatively low level of potential choice in Wales the matched English schools are located in largely rural areas. Hence, they conclude that 'It appears unlikely, therefore, that market-based accountability via parental choice is the main driver behind our results'. As a result we take this as evidence that the explanation for the difference in performance between England and Wales is due to the reputation model.

Discussion: behavioural analysis of the mechanisms behind the four models of governance

In this section we point to behavioural research which sheds light into the mechanisms whereby each of the four models can 'work'. We consider in order, altruism, hierarchy and targets, reputation, and choice and competition (which turns out to be especially problematic).

1. Altruism: the problem of inadequate feedback

For the purpose of this behavioural analysis, we reframe the 'altruism' model as 'the private provision of a report to the provider' (Hibbard et al, 2003). Why does private reporting improve performance? The answer is that private reporting is feedback and our human brain has evolved to learn and make good use of feedback. This is why '**giving feedback**' is one of the nudges described by Thaler and Sunstein (2008). Private feedback can improve performance because it *offers information about own performance*, and because it *increases salience of the criteria* used to measure performance. Because human attention is bounded by cognitive and motivational limitations, tools that attract attention and increase salience of important criteria are helpful when trying to change human behaviour. But the evidence we have presented suggest that such private reporting is, as argued by Berwick et al (2003), *insufficient feedback*. It could be made more effective if it were to include information about how to improve performance and were followed by a goal-setting plan (Kluger and DeNisi, 1998). Even with these additions, however, such feedback will often lack the necessary high powered incentives to counter-act inertia and generate the necessary drive to improve delivery of services.

2. Hierarchy and targets: the power of reference points and looming losses

We reframe the hierarchy and targets model as setting targets for providers with serious sanctions for failure and rewards for success. Prospect theory (Kahneman and Tversky 1979) explains why high powered incentives to improve the delivery of services are triggered by setting targets so that failure results in losses. The basic explanation is a strong ‘stability bias’ that anchors people to reference points (e.g., Samuelson and Zeckhauser 1988) and ‘loss aversion’: people feel losses more keenly than they do gains (Tversky and Kahneman, 1991). Hence the sanctions for failure are more powerful than rewards for success.

3. Reputation: the power of spotlight and moral emotions

We reframe the reputation model as ‘the provision to providers *and to the public* of regular reports that rank performance in ways that are easy *for the public* to understand’. A system designed to satisfy these criteria (Hibbard et al 2003) has been shown to have had the necessary high powered incentives to improve the delivery of services. From our evidence it looks to be the most effective of all models considered, at least in the short term. Why does reputation work? We offer two interpretations.

The first is the most discussed in the literature and revolves around the role of the *public* as the key trigger of change. If the public is aware of the existence of published performance data, and these data are a straightforward ranking of providers, this ranking sticks in the mind (Hibbard et al 2005). Even if users do not exercise choice, and market share is not itself affected, a poor ranking has an effect as published data make accountability relationships salient (e.g. Tetlock 1983). It is through this that poorly ranked providers are pressured to change and do change. Accountability encourages individuals to exert additional effort, so improves performance in tasks where additional effort helps, as for some decision errors (Simonson and Nye, 1992). In this interpretation, the reputation model exploits the same behavioural mechanism of loss aversion under Hierarchy and Targets.

We offer a second, more novel, interpretation which assumes no role of the public and only works via the mind of the provider from the shame, embarrassment and guilt as sequelae of being in the ‘spotlight’ (as shown by Thaler and Sunstein). There are a couple of psychological reasons why spotlight works. Slovic’s chapter describes one explanation: ‘affect’. Just like making a donation to a charity makes one ‘feel good’ (‘warm glow’, Kahneman and Knetsch 1992), receiving a low rank makes us ‘feel awful’. In this interpretation providers are not just ‘Econs’ responding to incentives, but ‘Humans’ who juggle – alongside economic considerations – their own emotions and ego (Kluger and De Nisi, 1998; Higgins 1997).

People generally like to think of themselves as moral and honest (e.g., Mazar and Ariely 2006), so why do providers need a public report to trigger quality improving exercises that they should have been doing all along? An answer in the most recent behavioural research is that just like judgments of probability, risk and value, also moral judgments are guided (or misguided) by the automatic and implicit ‘System 1’ of reasoning. As a consequence, we tend to use sub-optimal heuristics (moral

heuristics, Sunstein 2005) with predictable errors. On the upside, by mere virtue of being predictable, these errors can be corrected too by appropriate ‘moral design’ (Gigerenzer 2010). We argue that public reports could act as a particular instance of ‘moral design’ reminding public servants of why they chose to work in the public sector to start with. We have mentioned how public reports have the potential of shaking the public’s trust in the institution. Indeed, one of the most powerful moral heuristics ingrained in our mind is ‘do not betray’, which triggers the even more powerful reaction of ‘punish betrayal’. Reputation could work out of fear of having betrayed the public’s trust and provide an urgent reason for acting before the public reacts and ‘punishes’ this betrayal. This also means that systems that inflict reputational damage on failing organisations may be seen as objectionable. The problem is, however, that shocks of this kind are an integral part of generating the high powered incentives necessary for improvement.

4. Choice and competition: The problem of inadequate choice architecture

Behavioural research suggests that there are two fundamental reasons why the choice model might not have worked as well as expected. The first reason is inadequate implementation of the choice model. The second reason is more potent and involves inadequate architecture: the choice model makes the strong assumption that the public want to choose in situations where they cannot or do not want to exercise choice. Behavioural research provides a firm set of evidence-based guidelines on when provision of choice is expected to translate into exercise of choice, social welfare and subjective satisfaction (Botti and Iyengar, 2006) and none of them are met for health care. Some of these guidelines pertain again to format and amount of information, but the most important we review here concern the chooser and the context of choice:

- Individuals have articulated and stable preferences before the information is seen and the decision is made (Chernev 2003).
- Different individuals have different preferences and needs and options vary in the degree to which they meet these different tastes and needs.
- People have the knowledge and expertise, or the willingness and time to exert the effort to learn the relevant information, to make the choice (Loewenstein 1999).
- The choice is not among options with unpleasant outcomes and does not require trade-offs that cause psychological pain (Botti Iyengar 2004).

Applying these guidelines to school or hospital governance yields an interesting contrast⁹ and suggests that even if the choice model were well designed, it probably could only be effective for

⁹ In the case of secondary school choice, the choosers are parents who, by the time their children reach the age of 16, have a good idea of the type of educational setting they deem more suitable for their children (academic or sport-oriented?); children are different in their needs and tastes and – in most urban areas – there is a variety of school offerings that can cater to different needs and tastes; parents are typically involved in the decision and have multiple occasions for learning about the schools’ quality; and finally the choice revolves around giving opportunities for growth and future success of one’s offspring. In the case of hospital choice, the choosers are patients whose preferences for what is important to them is constructed in the process of choosing and is heavily influenced by the order in which information is presented; patients with a similar problem all have the same basic need to get the problem cured in the best possible manner and as soon as possible; patients perceive there is little variety among hospitals and feel it is almost unethical to expect that there is variety across hospitals; people do not have the medical knowledge, expertise and – being unwell – also little time or

school choice. Choice for hospitals remains a tricky context for applying a choice model, even when the choice is simplified or ‘nudges’ are used (such as opt-out defaults, feedback etc.). In the health context, people prefer ‘surrogate choice’ (i.e. choice by an advisor) to choice by themselves and to no-choice, which suggests that a softer choice model could be via the interplay of an ‘advisor’.

Conclusion: Policy, practice and theory

We have presented strong evidence that systems that are designed to inflict reputational damage, by satisfying Hibbard’s four criteria (Hibbard et al, 2003) have had powerful impacts in improving performance in what has been measured: for hospitals in the US, schools in the UK; and, alongside hierarchy and targets, for hospitals and ambulances in the UK. We have argued that an explanation for the power of the reputation model comes from seeing both users and providers of public services as ‘humans’ rather than ‘econs’. In England the policy preference for the NHS is choice and competition in a quasi market (Secretary of State for Health, 2010), which is paradoxical as recent evaluations of the policy introduced into the NHS have found only weak evidence for choice and competition in quasi markets being a sound lever for improvement (Audit Commission and Healthcare Commission, 2008; Brereton and Vasoodaven, 2010; Dixon et al 2010; Cooper et al, 2011; Gaynor et al, 2010). In England the policy preference for schools is both choice and competition and reputation (via published league tables). The other UK countries have abandoned school league tables, did not introduce analogues to ‘star ratings’ for their NHSs and are basing their policies on the assumption that those who are responsible for delivering services are driven by altruism.

Why do governments favour policies based on markets or altruism when these do not appear to be as effective as hierarchy, targets and reputation either in practice or in theory? This looks like a case of producer capture: the producers who suffer from ‘naming and shaming’ and hierarchy and targets influence those to whom they provide services and the press that such systems are iniquitous and unfair, so they become unpopular with the public. Ministers in England may be better able to ‘name and shame’ poorly-performing providers than Ministers in the smaller countries because of the greater ‘relational distance’ in England (Hood, 2007)¹⁰. But even in England, hospitals seem to have been able to capture the government: which continues to ‘name and shame’ ‘failing’ schools, but not ‘failing’ hospitals. This is paradoxical, because as we have argued, the choice and competition model appears to be more effective for schools than for hospitals. The explanation seems to be that providers of health care are more powerful than teachers and, as ever, politics trumps whatever evidence and theory might suggest.

References

willingness to learn about the choice; and finally the choice is among cures or places where there is a risk of dying, contracting serious infections and experiencing complications as well as the virtual certainty of experiencing discomfort, pain, separation from loved ones and interruption of normal daily routines (Boyce et al 2010).

¹⁰ Burgess et al (2010, p. 6) point out that the rationale for the Welsh Assembly government stopping the publication of school league tables was that they ‘do not have the support of either the teaching profession or members of the public’ and was part of a policy of ‘greater trust in producer determined solutions’.

- Audit Commission and Healthcare Commission (2008). *Is the Treatment Working?* London: Audit Commission, <http://www.audit-commission.gov.uk/reports/NATIONAL-REPORT.asp?CategoryID5andProdID59F8B7F6A-214D-4165-BE65-716315270A82andfromREPORTSANDDATA5NATIONAL-REPORT>
- Auditor General for Wales (2005) *NHS waiting times in Wales. Volume 1 – The Scale of the problem*. Cardiff: The Stationery Office. (Available from <http://www.wao.gov.uk/reportsandpublications/2005.asp>.)
- Berwick, D. M., James, B., and Coye, M. J. (2003). Connections Between Quality Measurement and Improvement. *Medical Care*, 41(1; sup), I-30-I-38.
- Besley, T. J., Bevan, G., and Burchardi, K. B. (2009). *Naming and shaming: the impacts of different regimes on hospital waiting times in England and Wales*. Centre for Economic Policy Research, Discussion Paper No. 7306,
- Bevan, G. (2010). Approaches and impacts of different systems of assessing hospital performance. *Journal of Comparative Policy Analysis*, 12 (1 and 2), 33 – 56.
- Bevan, G., and Cornwell, J. (2006). Structure and logic of regulation and governance of quality of health care: was OFSTED a model for the Commission for Health Improvement?. *Health Economics, Policy and Law*, 1 (4), 343-370.
- Bevan, G., and Hamblin, R. (2009). Hitting and missing targets by ambulance services for emergency calls: impacts of different systems of performance measurement within the UK. *Journal of the Royal Statistical Society (A)*, 172 (1), 1-30. <<http://www3.interscience.wiley.com/cgi-bin/fulltext/120848267/PDFSTART>>
- Bevan, G., and 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.
- Botti, S., and Iyengar, S. S. (2004). The Psychological Pleasure and Pain of Choosing: When People Prefer Choosing at the Cost of Subsequent Satisfaction. *Journal of Personality and Social Psychology*, 87 (3), 312–26.
- Botti, S., and Iyengar, S. S. (2006). The Dark Side of Choice: When Choice Impairs Social Welfare. *American Marketing Association*, 25 (1), 24-38.
- Boyce, T., Dixon, A., Fasolo, B., and Reutskaja, E. (2010). Choosing a high-quality hospital: the role of nudges, scorecard design and information. The King's Fund, London, UK. [http://www.kingsfund.org.uk/publications/choosing_a.html]
- Brereton, L., and Vasoodaven, V. (2010). *The Impact of the NHS Market*. London: Civitas, http://www.civitas.org.uk/nhs/download/Civitas_LiteratureReview_NHS_market_Feb10.pdf
- Burgess, S., Wilson, D., and Worth, J. (2010). *A natural experiment in school accountability: the impact of school performance information on pupil progress and sorting*. Bristol The Centre for Market and Public Organisation Working Paper No. 10/246.
- Chernev, A. (2003). When more is less and less is more: the role of ideal point availability and assortment in consumer choice. *Journal of Consumer Research*, 30, (2), 170–183.
- Chassin, M. R. (1996). Improving the quality of health care: what strategy works? *Bulletin New York Academy Medicine*, 73 (1), 81–91.

<<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2359372/pdf/bullnyacadmed01031-0089.pdf>>

- Chassin, M. (2002) Achieving and sustaining improved quality: Lessons from New York State and cardiac surgery. *Health Affairs*, 21 (4), 40-51.
- Connolly, S., Bevan, G., and Mays, N. (2010) *Funding and Performance of Healthcare Systems in the Four Countries of the UK before and after Devolution*. London: the Nuffield Trust.
- Cooper, Z., Gibbons, S., Jones, S., and McGuire, A. (2011). *Does Hospital Competition Save Lives? Evidence From The English NHS Patient Choice Reforms*. *Economic Journal*, 121, F228–F260.
- Coulter, A., N. L. Maistre and L. Henderson (2005) *Patients' Experience of Choosing Where to Undergo Surgical Treatment. Evaluation of London Patient Choice Scheme*. Oxford: Picker Institute Europe.
- Dixon, A., Robertson, R., Appleby, J., Burge, P., Devlin, N., and Magee, H. (2010) *Patient choice. How patients choose and how providers respond*. London: King's Fund.
- Dranove, D. K., McClellan, D. and Satterthwaite, M. (2003). Is more information better?: the effects of 'Report Cards' on health care providers. *Journal of Political Economy*, 111: 555–588.
- Fung, C. H., Lim, Y.-W., Mattke, S., Damberg, C., and Shekelle P. G. (2008). Systematic review: The evidence that publishing patient care performance data improves quality of care. *Annals of Internal Medicine*, 148, 111-123.
- Gaynor, M. S., Propper, C., and Moreno-Serra, R. (2010). *Death by Market Power: Reform, Competition and Patient Outcomes in the National Health Service*. Cambridge: NBER Working Paper No. w16164.
- Gigerenzer, G. (2010). Moral Satisficing: Rethinking Moral Behavior as Bounded Rationality. *Topics in Cognitive Science*, 2, 528-554.
- Hibbard, J. H. (2008). What can we say about the impact of public reporting? Inconsistent execution yields variable results. *Annals of Internal Medicine*, 148, 160 - 161.
- Hibbard, J.H., Stockard, J., and Tusler, M. (2003). Does publicizing hospital performance stimulate quality improvement efforts? *Health Affairs*, 22 (2), 84-94.
- Hibbard, J. H., Stockard, J., and Tusler, M. (2005) Hospital performance reports: impact on quality, market share, and reputation. *Health Affairs*, 24 (4), 1150-60.
- Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist*, 52, 1280-1300.
- 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(2): 95-102.
- Holmstrom, B. and Milgrom, P. (1991). Multi-task principal-agent analyses: Linear contracts, asset ownership and job design. *Journal of Law, Economics and Organisation*, 7, 24-52.
- Iyengar, S. S., and Lepper, M. (2000). When Choice is Demotivating: Can One Desire Too Much of a Good Thing? *Journal of Personality and Social Psychology*, 79, 995-1006
- Kahneman, D., and Knetsch, J. L. (1992). Valuing public goods: the purchase of moral satisfaction. *Journal of Environmental Economics and Management*, 22, 57-70.

- Kahneman, D., and Tversky A. (1979) Prospect Theory: An Analysis of Decision under Risk, *Econometrica*, XLVII (1979), 263-291.
- Kluger, A. N., and DeNisi, A. (1998). Feedback Interventions: Toward the Understanding of a Double-Edged Sword. *Current Directions in Psychological Science*, 7 (3), 67-72.
- Le Grand, J., Mays, N., and Mulligan, J. (eds). (1998) *Learning From the NHS Internal Market*, London: King's Fund.
- Le Grand, J. (2003). *Motivation, agency and public policy: Of knights and knaves, pawns and queens*. Oxford: Oxford University Press.
- Le Grand, J. (2007). *The Other Invisible Hand: Delivering Public Services Through Choice and Competition*. Woodstock: Princeton University Press.
- Lowenstein, G. (1999). Is More Choice always Better? *Social Security Brief*, 7, 1-7.
- Mannion, R., Davies, H. and Marshall, M. (2005a) Impact of star performance ratings in English acute hospital organisations. *Journal of Health Services Research and Policy*, 10, 18–24.
- Mannion, R., Davies, H. and Marshall, M. (2005b) *Cultures for Performance in Health Care*. Maidenhead: McGraw Hill.
- Marshall, M. N., Shekelle, P. G., Leatherman, S., and Brook, R. H. (2000). The Public release of performance data: What do we expect to gain? A review of the evidence. *Journal of the American Medical Association*, 283 (14), 1866-1874.
- Mazar N. and Ariely D. 2006; Dishonesty in Everyday Life and its Policy Implications. *Journal of Public Policy and Marketing*. Vol. 25, No. 1: 117-126
- Propper, C., Sutton, M., Whitnall, C. and Windmeijer, F. (2008). Did 'targets and terror' reduce waiting times in England for hospital care? *The B.E. Journal of Economic Analysis & Policy*, 8, no. 2, article 5. (Available from <http://www.bepress.com/bejeap/vol8/iss2/art5>.)
- Propper, C Sutton, M., Whitnall, C. and Windmeijer, F. (2010). Incentives and targets in hospital care: Evidence from a natural experiment. *Journal of public economics*, 94 (3-4), 318-335.
- Secretary of State for Health. (2010) *Equity and excellence: Liberating the NHS*. Cm7881. London: Department of Health.
<http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_117353>
- Simonson, T., and Nye, P. (1992). The effect of accountability on susceptibility to decision errors. *Organisational Behavior and Human Decision Processes*, 51 (3), 416-466.
- Sunstein, C. R. (2005). Moral Heuristics. *Behavioral and Brain Sciences*, 28 (4), 531-542.
- Teelken, C. (2000). Market forces in education, a comparative perspective in England and Scotland, *Scottish Educational Review*, 32 (1), 142-155.
- Tetlock, P. E. (1983). Accountability and complexity in thought. *Journal of Personality and Social Psychology: Attitudes and Social Cognition*, 45, 74-83.

Thaler, R. H., and Sunstein, C. R. (2008). *Nudge*, London: Yale University Press.

Tversky, A., and Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Economics*, 106, 1039-1061.

Numbers waiting for elective hospital admission ('000s)

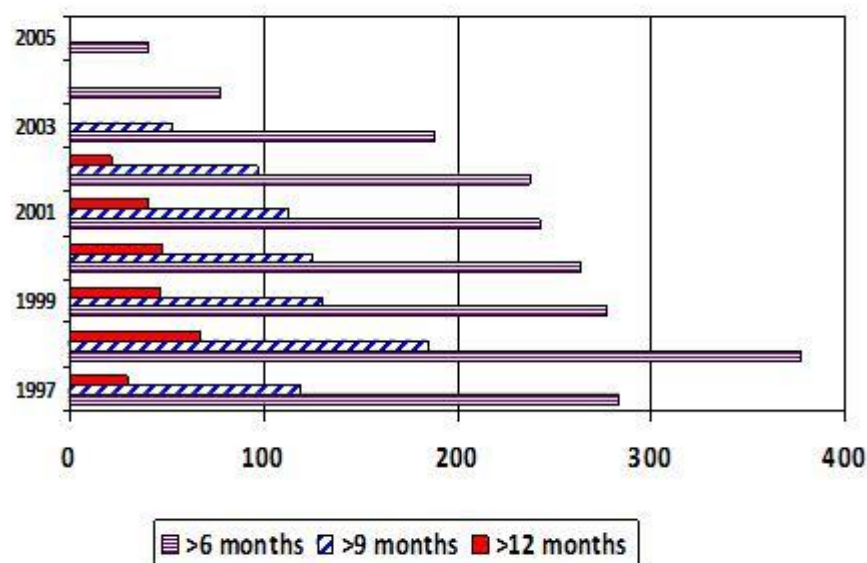
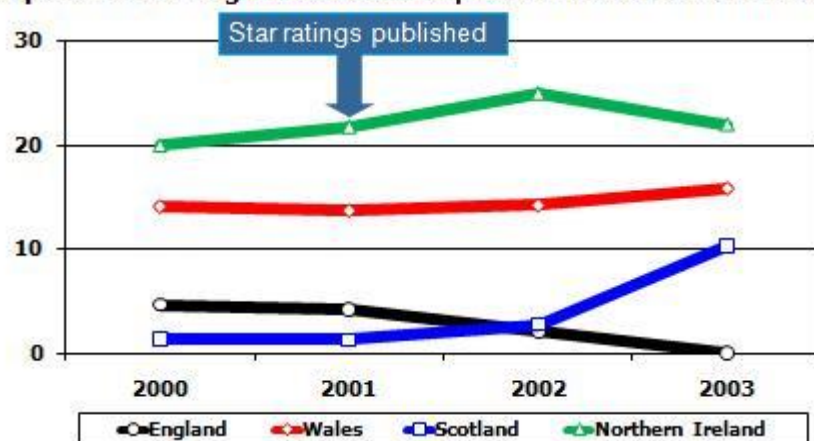


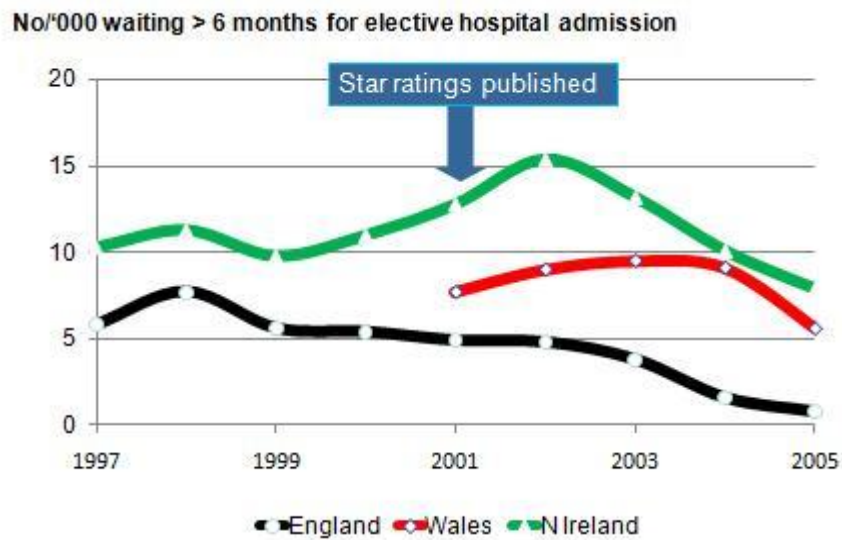
Figure 1: numbers waiting more than 6, 9 and 12 months for elective hospital admission in England from 1997 to 2005

% patients waiting for elective hospital admission > 12 months



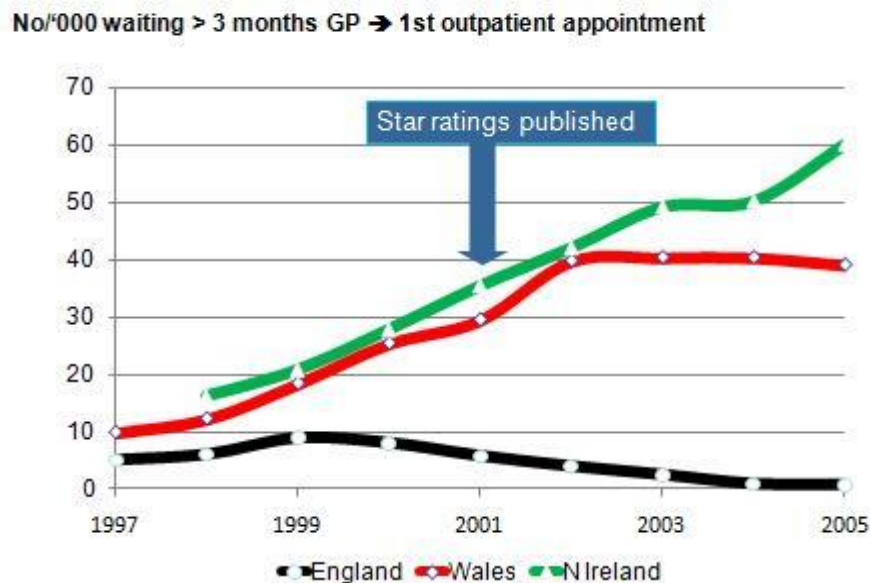
Source: <http://www.statistics.gov.uk> National Health Service hospital waiting lists by region: Regional Trends 35, 36, 37 & 38

Figure 2: percentage of patients on waiting lists for hospital admission waiting more than 12 months in England, Wales, Scotland and Northern Ireland from 2000 to 2003



Source: Bevan (2009)

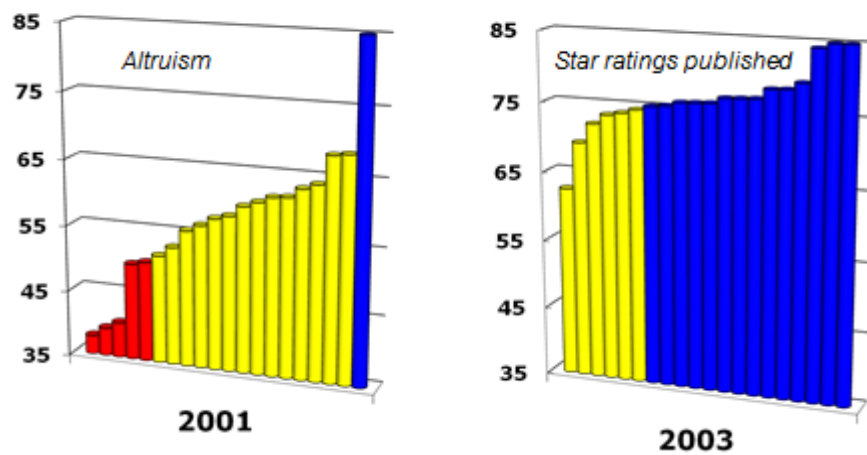
Figure 3: numbers per thousand waiting more than 6 months for elective hospital admission in England, Wales, and Northern Ireland from 1997 to 2005



Source: Bevan (2009)

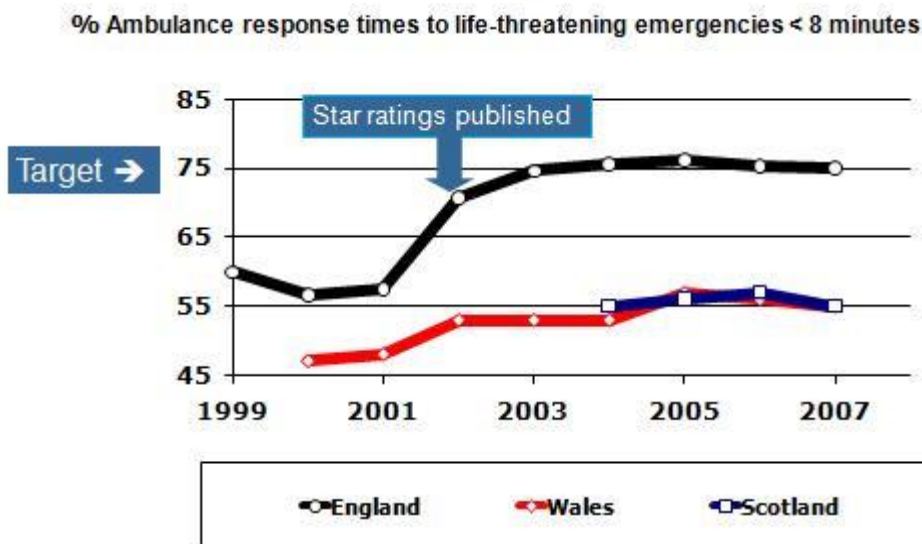
Figure 4: numbers per thousand waiting more than 3 months for GP referral to a specialist in England, Wales, and Northern Ireland from 1997 to 2005

% Ambulance response times to life-threatening emergencies < 8 minutes
(Target 75%)



Source: Bevan & Hamblin (2009)

Figure 5: percentages of ambulance responses within eight minutes to Category A calls by ambulance services in England in 2001 and 2003



Source: Bevan & Hamblin (2009)

Figure 6: national average percentages of ambulance responses within eight minutes to Category A calls for England, Wales and Scotland from 1999 to 2007

Pupils / teacher public secondary schools

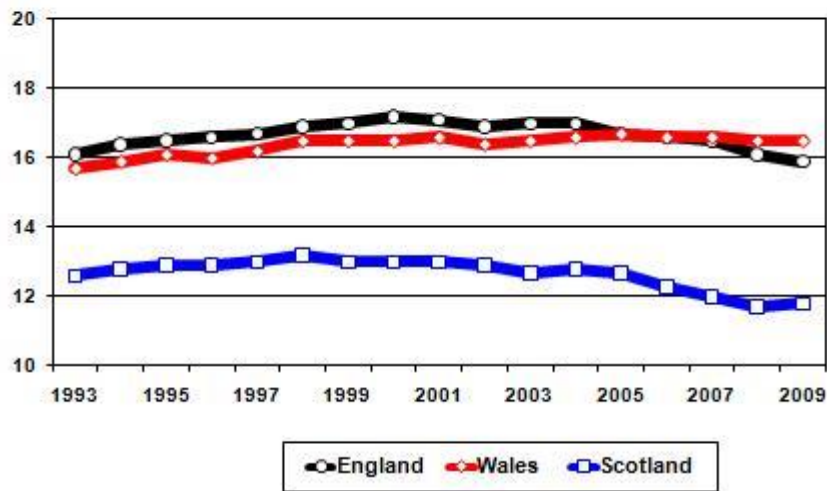
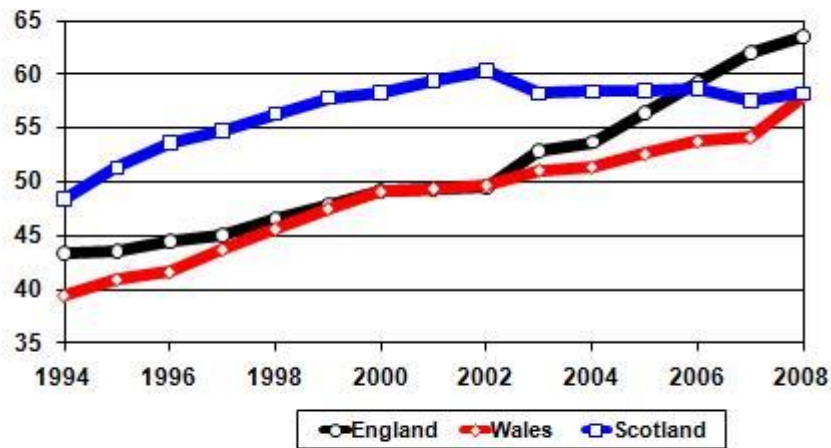


Figure 7: pupil teacher ratios in secondary schools in England, Wales and Scotland from 1993 to 2008

% pupils achieving > 5 good grades at age 16*



Note: data are missing for 1998, 2001 & 2005 & these have been estimated as the mean values for adjacent years

Figure 8: examination performance in terms of five good grades at age 16 in England, Wales and Scotland from 1994 to 2008

Box 1: Targets for Waiting Times in England and Wales

Service	England (2004 and 2005)	Wales (end March 2005)
First outpatient appointment	March 2004: 17 weeks December 2005: 13 weeks	18 months.
Inpatient/day case treatment	March 2004: 9 months December 2005: 6 months.	18 months*
Potential longest overall waiting time within current national target for outpatients and inpatient/day cases	March 2004: 13 months December 2005: 9 months.	36 months.

Source: Auditor General for Wales, 2005a, NHS waiting times in Wales. Volume 1 – The Scale of the problem (Cardiff: The Stationery Office), p. 16.
<http://www.wao.gov.uk/reportsandpublications/2005.asp>

* With a guarantee of an offer of alternative treatment for waits over twelve months by 31 March 2005

Box 2: similarities in policies for schools and hospitals in England in the 1980s and 1990s

	Hospitals in England	Schools in England
Quasi markets	From 1991 to 1997: 'money followed the patient' and hospitals were separated from health authority control.	From 1988: 'money followed the pupil' and allowed schools to opt out of local authority control.
National standards	From 1997: National Institute of Clinical Excellence (NICE) to develop guidelines for cost-effective care and appointed 'Czars' to develop standards for priority services (such as cancer and coronary heart disease) through national service.	From 1996 national curriculum (Education Act, 1996, Chapter 56, sections 358 – 63) frameworks
Creation of new inspectorates to inspect all organisations over a four-year period	1999 - 2004: Commission for Health Improvement (CHI) Commission for Health Improvement (CHI)	From 1992: the Office for Standards in Education (OFSTED)
Publication of comparative performance ranked on a national basis	2001 to 2005: annual 'star ratings'	From 1994: annual league tables for secondary schools (see e.g., Department for Education and Skills, 2006).

Box 3: comparisons of policies for schools in England, Wales and Scotland

	England	Wales	Scotland
Opting out	From 1988		From 1989 but with limited effects as compared with England
Local Management	From 1988 to governing bodies		From 1993 to headteachers
Quasi markets	From 1988		Choice in principle but weak marketisation as compared with England
National curriculum	From 1988		None
Inspections with reports published	From 1992 by OFSTED	From 1992 by OFSTED and then from 1998 by IESTYN*	By Her Majesty's Inspectorate of Education (which dates back to 1840) from 1998 was made an Executive Agency of the Scottish Government
School league tables published ranking performance	From 1994	From 1994 to 2001	From 1998 to 2002

* IESTYN is Her Majesty's Inspectorate for Education and Training in Wales