Targets, Choice and Voice: Accountability in Public Services

Deborah Wilson

2020 Public Services Trust *at the* RSA

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

Targets, Choice and Voice: Accountability in Public Services

Deborah Wilson

2020 Public Services Trust
at the RSA

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

About the
2020 Public Services Trust

The 2020 Public Services Trust is a registered charity (no. 1124095), based at the RSA. It is not aligned with any political party and operates with independence and impartiality. The Trust exists to stimulate deeper understanding of the challenges facing public services in the medium term. Through research, inquiry and discourse, it aims to develop rigorous and practical solutions, capable of sustaining support across all political parties.

In December 2008, the Trust launched a major new **Commission on 2020 Public Services**, chaired by Sir Andrew Foster, to recommend the characteristics of a new public services settlement appropriate for the future needs and aspirations of citizens, and the best practical arrangements for its implementation.

For more information on the Trust and its Commission, please visit www.2020pst.org.

The views expressed in this report are those of the author and do not represent the opinion of the Trust or the Commission.

About the ESRC

The Economic and Social Research Council is the UK's leading research and training agency addressing economic and social concerns. We aim to provide high-quality research on issues of importance to business, the public sector and Government. The issues considered include economic competitiveness, the effectiveness of public services and policy, and our quality of life. The ESRC is an independent organisation, established by Royal Charter in 1965, and funded mainly by Government. Economic and Social Research Council Polaris House North Star Avenue Swindon SN2 1UJ

Telephone: 01793 413000
www.esrcsocietytoday.ac.uk

About the Author

Deborah Wilson is a Senior Research Fellow at the Centre for Market and Public Organisation at the University of Bristol and Deputy Director of the ESRC Public Services Programme. She can be contacted at:

Deborah Wilson
CMPO
University of Bristol
2 Priory Road
Bristol BS8 1TX
d.wilson@bristol.ac.uk

Acknowledgements

Introduction to this series

The Commission on 2020 Public Services is a major inquiry into how public services should respond to the significant societal challenges of the next decade. The Commission is developing a practical but compelling vision of the priorities for public action to address the emergent challenges facing society in 2020. The Commission has three aims:

1 To broaden the terms of the debate about the future of public services in the UK.
2 To articulate a positive and long-term vision for public services.
3 To build a coalition for change.

This series of essays represents a working partnership between the 2020 Commission and the Economic and Social Research Council (ESRC). As part of our commitment to rigorous, evidence-based research, we jointly commissioned a series of experts to examine the key issues in public services. Two broad themes emerged: one considering future relationships between citizens, state and society; the other exploring the future delivery of public services.

Generous support from the ESRC has allowed the Commission to dig deep into a complex set of issues, and ensure its inquiry represents the best contemporary thinking on public services and society, with a strong evidence base.

Each paper can be read separately, and will also be available as a collected volume in the future. We believe that the research and analysis emerging from this partnership is a rich and significant contribution both to the ongoing national debate on public services and to the Commission's vision for the future. We hope that you enjoy the series, and we invite you to share your own reflections and analysis at www.2020pst.org.

Foreword

"Every area in the world you go is different… so different circumstances will be there and different things will be wanted by the people living there." Male, 25-39, Oxford

"Let's focus on national standards, you've still got to have a standard of some sort to deviate from or not." Male, 40-64, Kent

The future of public services is the number one issue for the new coalition government. After an election campaign in which all three major parties skirted nicely around the biggest challenges ahead, the country must now prepare for an emergency budget and a spending review that will begin to fundamentally re-shape the way our public services are configured.

Decisions must be made about what kind of public services we can afford, who should pay for them, and what public priorities should be at a time of fiscal parsimony. The 2020 Public Services Trust has argued that these decisions should be made both strategically and democratically - with a long-term vision and the needs, wants and aspirations of future citizens at the centre. How we account for the performance of public services is central to this strategic approach.

In its interim report Beyond Beveridge, the Commission on 2020 Public Services argued for three shifts:

- A shift in culture – from social security to social productivity. This means moving away from thinking about public services in terms of inputs and access, and towards a focus on outcomes and relationships.
- A shift in power – from the centre to citizens. This means moving away from an over-centralised central government machinery, and towards building public services around people and places.
- A shift in finance – reconnecting finance to purpose. This means that the

financing of public services should reinforce their purpose, aligning incentives and resources better through people's lives, and fostering a more open and transparent fiscal model.

Against all of these arguments, questions of what to measure, how to measure it, and who should be held accountable are key. Yet the two quotes above - from participants in deliberative group events the Commission held earlier this year - illustrate some of the tensions at play here. How much flexibility should localities, departments or clinicians have over spending decisions? How much 'difference' can a national system allow? How can potentially life-changing decisions be made most effectively and democratically? And what is the proper balance between central control and citizen power?

The following two papers by Dr Deborah Wilson begin to provide some answers to these questions. In the first, Dr Wilson explores some key debates around targets- and rankings-based accountability mechanisms, unpicking some of the trends that have driven an increasingly 'tight' and centralised system. Her second paper addresses citizen control - asking how user-based mechanisms such as choice and voice can drive public service performance and responsiveness from the bottom up.

Dr Wilson's accounts show that there are limits to the effectiveness of both top-down and bottom-up accountability mechanisms. The problems of measurement error and gaming undermine the effectiveness of performance indicators. Meanwhile the reality of bottom-up accountability has rarely matched the rhetoric, and is very contextually dependent. What is further evident from these papers is the extent to which the two forms of accountability are interdependent, and the subsequent importance of achieving consistency between the incentives created by them.

The two excellent papers in this publication are part of a major partnership between the ESRC and the 2020 Public Services Trust, in which a range of experts have thought creatively about the present and the future of public services. In an era of spending cuts and efficiency savings, it is imperative that we understand how the quality and responsiveness of public services can be accounted for most effectively. These papers provide a means to better understand this issue and should form a vital contribution to the policy debate.

Henry Kippin & Lauren Cumming
2020 Public Services Trust, June 2010

1
Targets- and rankings-based accountability mechanisms

Abstract

Targets, rankings and other forms of 'managing by numbers' are now a familiar feature of public service management. Performance indicators (PIs) are widely used as part of broader accountability mechanisms and/or incentive schemes aimed at improving public service performance. For example, targets may be accompanied by explicit financial incentives: rewards for hitting the target, sanctions for missing it. Or some form of market mechanism may provide an implicit incentive structure, with an organisation's league table position affecting user choice and in turn the financial stability of the organisation, for example. Such incentives can be directed at the level of the individual as well as the organisation, and public service providers often face elements of a range of different accountability mechanisms concurrently.

In this paper I review the evidence on the effects of 'managing by numbers' across health and education. I first describe the range of ways in which performance indicators may be employed, and discuss the particular issues associated with their use in the public sector and the range of responses that may result. Second, I explain the different types of PI – raw outcomes, 'value added', composite measures – and discuss the limitations of each. I then review the evidence on whether and how public service providers respond to such indicators, and show how that response may be determined both by the type of PI employed and the broader accountability mechanism of which it is a part. Providers do respond to the incentives created by such schemes, but this may not be in ways that are anticipated or desirable. Finally, in light of the evidence, I consider the lessons learnt about the use of PIs and the subsequent implications for their future use.

Introduction

Targets, benchmarks and league tables have a long history in both the private and public sectors. While such metrics are now a familiar feature of public service management, the principle of measuring performance relative to some published measure goes back at least to Jeremy Bentham's plans for prison management in the 1790s (Bevan and Hood 2006a). Their use in the UK has accelerated since Tony Blair came to office. His administration introduced more than 300 performance targets across all government departments in 1998, directly linked to budget allocations and translated into many more targets and indicators at lower levels of government (Hood 2006). Since devolution, however, it is England rather than its neighbours that continues to emphasise the use of PIs in its public services performance regimes.

At least three factors seem to be behind the increasing reliance on the use of performance indicators (PIs) (Hood et al 2009). First, the long term development of a performance measurement movement, spanning the worlds of consultancy, academia and government agencies, and aided by advances in IT. Second, the appeal to some public managers of apparently 'transparent' steering processes such as rating and ranking, as opposed to traditional, more opaque, methods. Third, the appeal to politicians of seemingly 'objective' systems that do not rely on trust in public service providers; that provide demonstrable evidence of improvement; but that can still be finessed to suit political needs through choice of, and changes in, the PIs themselves.

Notwithstanding their widespread use, there are well-documented problems with using metrics-based performance management regimes in the delivery of public services (Propper and Wilson 2003; Bevan and Hood 2006a). There is also a growing body of evidence on the shortcomings of both the PIs themselves and on the responses of individuals and organisations to their use in alternative accountability mechanisms (Wilson and Piebalga 2008). Conversely, there is still much less evidence on whether or not PIs achieve the aim of improving the quality of public service delivery, and very little discussion regarding the relative costs of achieving any such improvement (Propper and Wilson 2003). Public service outcomes are complex and multi-dimensional. Measurement of improvements in public service quality is therefore difficult, as is attributing their cause to any one specific element of a more general performance management regime. I discuss all these issues below.

The aim of this paper is to review the evidence on the effects of targets and rankings, i.e. of 'top-down' performance management regimes, in public service delivery. In a companion paper (Wilson 2010a) I discuss 'bottom-up' accountability mechanisms such as choice and voice. I focus on the domains of health and education, and predominantly consider evidence from the US and the UK. These countries are seen as leaders in the use of PIs in these sectors and, as a consequence, the literature in this field has a predominantly UK and US focus. I first describe the ways in which PIs may be employed, and discuss the particular issues associated with their use in the public sector and the (potentially undesirable) responses that may result. I then explain the different types of PI and discuss the limitations of each. Drawing on my review of the evidence of their effects, I consider the implications for the use of PIs in the future.

What is 'managing by numbers'?

Governance systems using quantifiable performance indicators such as targets are described by Bevan and Hood (2006, 518-9) as a form of control "in which: (1) desired results are specified in advance in measurable form; (2) some system of monitoring measures performance against that specification; and (3) feedback mechanisms are linked to measured performance." Both (2) and (3), the system of monitoring and the feedback mechanism, can be designed in various, non-mutually exclusive, ways (Burgess et al 2002; Propper and Wilson 2003).

The measured performance information may be kept internal to the organisation or it may be published. If kept internal, it is essentially a management tool. There has been, however, an increasing trend for the publication of performance information which, at least according to the current administration's plans, is set to continue (HM Government 2009). If the information is made public, then there is a range of possible feedback mechanisms. The information may be used as part of an implicit, or indirect, incentive scheme, under which the organisation gets a financial reward as a result of the response of others to the information. A classic example of this is a 'quasi-market', in which providers of services are rewarded for good performance by getting more contracts. More generally, such implicit incentive schemes relate to 'bottom up' accountability mechanisms, in which the primary role of quantifiable performance information is to inform and empower consumers who then lever improvement in services through actions such as choice and/or voice. Such mechanisms are the focus of the companion paper to the present one

(Wilson 2010a), although there are of course links between the two (see also Allen and Burgess 2010).

In this paper the focus is on the use of published performance information as part of 'top down' incentive schemes, which explicitly specify how measured performance, often relative to a pre-specified target or to the performance of other provider organisations, is linked to sanctions and/or rewards. These sanctions/ rewards may be financial or non-financial, and can be targeted at the level of the individual, the team, the organisation or the policy area. So, for example, hospital managers in England faced dismissal if their hospital performed poorly against waiting times targets; conversely good performers were granted earned autonomy and the freedom to keep certain surpluses (Propper et al 2008a). The Quality and Outcomes Framework introduced in 2004 gave family doctors in the UK direct financial incentives to hit targets in the provision of a wide range of specified primary care treatments (Hood et al 2009). Alternatively, it may be career concerns or reputational effects, with implications that are played out via the labour market, that create the incentive for effort (Burgess and Ratto 2003). Team-based financial incentives were introduced within Her Majesty's Customs and Excise department (since merged with the Inland Revenue) as a result of the Makinson report of 2000 (Burgess et al 2009). A key point to note is that public service providers often face elements of a range of different accountability mechanisms concurrently (Kane and Staiger 2002). This raises problems of both attributing effects to any one mechanism and of minimising potential conflicts between the incentives created by the different structures (Wilson 2009).

There are two, linked, assumptions underlying the theory of governance by numbers (Bevan and Hood 2006a). First, measurement problems are not important: the inevitably incomplete measure of organisational performance contained within a quantifiable performance indicator (PI) adequately represents overall performance. Second, while it changes organisational and/or individual behaviour, such a governance system is not vulnerable to gaming. The particular features associated with the delivery of public services mean that neither assumption is likely to be valid.

Dixit (2002) stresses two important features of the public sector. The first is that bureaucrats often serve several stakeholders, including the users of the service, payers for the service and politicians. The second, a consequence of the first, is that the organisation and, therefore, the bureaucrats who work in it, often have several ends to achieve. These features are termed multiple principals and multiple tasks

respectively, and result in a public agency facing multiple, sometimes conflicting, goals. Wilson et al (2006), for example, show how headteachers face potential conflicts in their attempts to satisfy the demands of pupils, parents, and local and central government officials.

These multiple goals and often complex, multidimensional, outputs of public service organisations mean that performance relative to these goals is difficult to measure: any quantifiable representation of that output is therefore necessarily incomplete. This creates the opportunity for gaming, the extent of which may depend on individual motivation. Individuals will respond to PIs in ways that maximise their own utility or benefit. This is not necessarily consistent with PIs improving welfare, nor is it necessarily in ways that are expected by those who design the system. In Le Grand's (2003) terminology, responses will differ depending on whether providers are 'knights' or 'knaves'. Moreover, provider motivation may be endogenous to the type of performance management scheme imposed: differently motivated individuals may be attracted to alternative cultures of performance management (Gregg et al 2008). This relates to the literature on whether public service workers exhibit 'intrinsic' motivation (Frey 2000; Crewson 1997; Deci 1971) and how that may affect their responses to different forms of governance structure.

So neither assumption is likely to hold: managing by numbers in the public services will suffer from both measurement error and the potential for undesired as well as desired responses. The precise way(s) in which these combine with individuals' (unobserved) motivation to create the overall outcomes may in practice be difficult to disentangle (Bevan and Hood 2006a). Before reviewing the evidence on the effects of managing by numbers, therefore, I first discuss the forms that gaming behaviour may take and then look at the incomplete measurement issues arising from the use of different forms of PI.

Gaming responses

Smith (1995) provides a list of potential unintended consequences of publishing PIs in the public sector. These include tunnel vision, myopia, measure fixation, sub-optimization, gaming, misrepresentation and misinterpretation, and can be as unpredictable as schools boosting the calorific intake of their menus on days on which tests are held to boost student performance (Figlio and Winicki 2005). Hood (2006; 2007) discusses the different forms of strategic or gaming behaviour that result from target and ranking systems. Target systems produce the potential for

ratchet effects, threshold effects and output distortions. Ratchet effects arise in systems where the target is set as an incremental advance to 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.

Ranking systems (league tables) may also create similar incentives for output distortion. If the measured performance of an organisation is being directly – and publicly – compared to that of others, the incentive is to focus effort on that measured element in order to boost the ranking position. The extent to which this deflects from effort towards true improvement will depend on the extent to which the PI accurately reflects overall performance. The multi-product, or multi-task nature of many public service organisations' output gives agents opportunities to divert activity away from non-incentivised tasks, which furthers the potential for output distortions (Propper et al 2008b). There are examples of all such responses below.

Forms of PI

The simplest form of PI are raw outcomes, or *levels* (Kane and Staiger 2002), which measure the outcomes of an organisation or programme at some designated date: the number of individuals who do not die after emergency admissions for heart attacks, for example, or the number of pupils passing examinations at a certain grade. While they are easy to understand and relatively low cost to collect, they only deal with one dimension of a complex output, and do not provide sufficient information to isolate the impact of the organisation on the measured outcome. Patients may have recovered from heart attacks without medical intervention, for example. They are also susceptible to gaming: by adjusting the quality of the intake, an organisation can boost its performance as measured by raw outcomes. Such cream skimming practices are well documented in health and in education and are discussed below. *Levels* PIs may unfairly penalise effective providers serving disadvantaged and high cost populations while at the same time hiding poor performance from those serving lower cost populations. Risk adjustment of such measures helps account for heterogeneity in the populations served and therefore helps to both give a better measure of the impact of the agency and reduce the incentive for cream skimming (Propper and Wilson 2003; 2006).

Further to such risk adjustment, the aim of value-added or *gains* PIs (Kane and Staiger 2002) is to better isolate the impact of, for example, the school environment on pupil progress between two points in time. Such PIs do this by incorporating prior attainment and/or other factors which are outside the school's control but which are known to impact on test score (i.e. raw) outcomes (see Wilson and Piebalga (2008) for a detailed discussion of the range of value-added PIs that have been published in the English school league tables). While both reducing the incentive to cream-skim by explicitly accounting for input, and providing a better measure of the effectiveness of an organisation, *gains* PIs still generally only reflect one dimension of an organisation's output: in the case of schools, the focus is still on test scores (or progress between tests at different stages of schooling).

One response to this aspect of measurement error has been the development and increased use of composite indicators that attempt to combine many dimensions of an organisation's output into a single figure or rating. Examples include the star rating systems for hospitals in England and the CPA (comprehensive performance assessment) system of rating local government. While intuitively appealing and easy to understand, they are in practice complex and opaque (Hood 2006) and create alternative, and various, sources of measurement error.

The star rating system, for example, has been shown to be extremely sensitive to aggregation methods, either altering weightings or decision rules (Jacobs and Goddard 2007). Studies have also found that there is no relationship between hospitals' star ratings and either the organisations' productivity (Stevens et al 2006) or the clinical quality of adult critical care provided (Rowan et al 2004). McLean et al (2007) highlight various problems with the CPA system, including the sensitivity of CPA scores to the measures of deprivation employed. If PIs are unreliable in ways such as these, the public cannot be sure that performance is being meaningfully measured, and public service providers have no firm basis for assessing their performance (Hood et al 2009). It should also be noted that such composite rating systems have tended to be unstable, with changing methodologies and limited lifespans, which makes it difficult to make meaningful comparisons of organisational performance over time. Both CPA and star ratings have now been replaced with alternative performance regimes, for example.

Two further points are worth emphasising here. First, different aspects of a public service organisation's performance will matter more or less to its different stakeholders. Both the designers and final users care about a vector of outcomes, but the weights

these two principals attach to these outcomes may differ (Propper and Wilson 2003). This has implications for the aggregation methods employed for calculating composite indicators. Or consider a value-added measure of school performance: if it is <u>not</u> corrected for school resources (as is currently the case with the English contextual value added (CVA) measure) it is more suited to informing parental choice than to the government aim of raising standards for the same resource base (raising efficiency). The same PI may not be able to satisfy the often multiple purposes for which it is employed or the different concerns of multiple stakeholders.

Second, we need to distinguish the shortcomings of the PIs themselves from the shortcomings of the uses to which they are put. For PIs to be used as part of meaningful ranking exercises they need to discriminate clearly between different units. While the CVA measure for English schools contains useful performance management information, for example, it is not possible to use it to meaningfully rank individual schools. Once the uncertainties associated with its calculation are taken into account, Wilson and Piebalga (2008) find that over half of all English secondary schools are not significantly different from the national average when 'ranked' on their CVA scores. Jacobs et al (2007) obtain similar findings for hospital star ratings.

How public service providers respond to PIs

There is now a large and growing body of evidence on how individuals and organisations respond to 'managing by numbers' and I review this here. There is much less evidence on whether PIs actually achieve the aim of improving public service outcomes, although this is ideally the question that should be asked. There is even less evidence on the costs of achieving any such improvement, and I return to this point below.

Assessment of the impact of PIs is hampered by the lack of experimentation in, and associated assessment of, government policy (Propper and Wilson 2003). PIs have been introduced, generally, not in a controlled trial manner, but as a result of a policy change. Experiments may be viewed as unethical or too expensive. Often, performance assessment is accompanied by changes in other incentives. For example, in the UK, school league tables were introduced across all schools as part of the general reform of schooling provision. This makes it difficult to isolate the impact of the introduction of PIs from other policy changes that are implemented at the same time, including increases in resource levels (Hood 2006; Burgess

et al 2002). Differences in policy across the devolved administrations of the UK have created the potential to exploit natural experiments and thus better isolate the impact of specific policy changes; I report the results from one such study below. Finally, remember that improvements in measured performance does not necessarily imply real performance improvement, and/or may come at the expense of a distortion of effort away from other (unmeasured) aspects of public service delivery (Hood 2006). All these responses, and combinations thereof, have been found in each of the two sectors I now review.

Education

Education is one sector which has been subjected over many years to a variety of explicit and implicit incentive structures, often concurrently, which employ a range of different forms of PIs, in the UK, the US and elsewhere (Kane and Staiger 2002; Ladd 2002; Wilson 2004; Wilson and Piebalga 2008). It is therefore difficult to isolate the impact of any one form of PI on test score improvements; harder still to assess their effects on a broader definition of educational outcomes (Propper and Wilson 2003; Allen and Burgess 2010).

Early evidence on the impact of metrics-based accountability systems on student achievement in the US is described as 'limited and ambiguous' by Kane and Staiger (2002: 106), who express caution about attributing observed gains to the specific accountability systems employed in Texas and North Carolina. Clotfelter and Ladd (1996) are similarly cautious about the introduction of the Dallas accountability system: Dallas test scores did rise relative to other Texas cities, but the timing of this improvement predated the accountability system by a year. Ladd (1999) extends this analysis of the impact of the Dallas programme and finds some evidence of it having a positive impact on student outcomes: there are positive and relatively large effects for White and Hispanic seventh-graders, but not for their Black peers. Similarly Dee and Jacob (2009) find differential effects: positive effects on maths scores but no impact on reading scores. Jacob (2002) shows that test scores in maths and reading both increased in Chicago public schools after the introduction of a test-based accountability policy, but that student effort and improvements in test-specific skills largely drove those improvements. Hanushek and Raymond (2005) find a positive effect of such an accountability policy, which is only significant where sanctions are imposed for poor performance. Simply publishing league tables has no effect on school performance.

Attributing the observed gains in test score outcomes in England to school league tables is difficult, as they were introduced across all schools and as part of a much broader programme of education reform (Propper and Wilson 2003). It is also not clear that there has been sustained performance improvement. In the case of literacy and numeracy targets for English primary schools, there is some evidence of a one-off rise followed by flatlining (Hood et al 2009) although Tymms (2004) argues that both the National Curriculum and the form of the tests themselves have both evolved to such an extent that it is very difficult to assess what changes in test scores mean. Bradley et al (2000) do find evidence that the publication of PIs (allied with implicit incentives via the quasi-market) has produced an improvement of outcomes as measured by the PIs themselves. This is accompanied, however, by some evidence of cream-skimming.

More generally, there is a fairly large body of evidence on the potential undesired responses to league tables and targets in education. Here I highlight three manifestations relating to the types of gaming response discussed above. The first follows from the fact that any PI is inevitably an imperfect measure of a complex process, which creates the incentive to focus on those parts of the process that are measured, possibly to the detriment of other, less quantifiable, tasks. There are various ways in which such output distortions have been seen in the education sector. Wiggins and Tymms (2002) provide evidence of this 'narrowing' effect on the curriculum at primary level in the UK; Deere and Strayer (2001) show that passing rates on the tests included in the Texas school accountability system have increased relative to other (not included) tests; and Jacob (2002) shows how teachers responded to the test-based accountability system in Chicago public schools along several dimensions, one of which was substitution away from low-stakes subjects, such as science and social studies.

Second, for the same quality of education received, the better the input (the higher the ability of the pupils) the better the output and hence the higher the school's relative position in a levels-based ranking exercise. A *levels* PI therefore creates the incentive for cream-skimming. There is evidence of schools undertaking various cream-skimming strategies in response to such PIs (Meyer 1997). Figlio and Getzler (2002) and Cullen and Reback (2002) both provide evidence of US schools reclassifying weak students in order that they are not eligible for the tests that are the subject of the indicator. Schools may additionally have the incentive to engage in cream skimming at the point of admission (Gewirtz et al 1995; Whitty et

al 1998; West and Pennell 2000). Value added PIs, which attempt to better account for heterogeneity in the pupil population, is one possible response to the problem of cream skimming.

Finally, target PIs create threshold effects which may impact on within-school resource distribution. They introduce an essentially arbitrary dichotomy into continuous data and therefore focus attention on the borderline (Fitz-Gibbon and Tymms 2002). We may expect schools to shift their activities or target their resources to pupils who are expected just to miss the target in the absence of (extra) intervention. This may be to the detriment of pupils at either end of the ability distribution, and may or may not be welfare improving. There is evidence of this behaviour both in the US and the UK. For the US, Deere and Strayer (2001) show that it is those pupils at or below the passing level that exhibit the most improvement in the tests included in the Texas accountability system. Neal and Schanzenbach (2007) find highly significant gains in the middle deciles of the ability (prior achievement) distribution, in contrast with more limited gains at either end. Reback (2008) explicitly calculates the estimated probability of pupils and their classmates affecting their school's rating based on their prior attainment, and finds that a pupil's marginal importance to their school's rating is associated with faster test score improvement.

For the UK, Wiggins and Tymms (2002) provide evidence of such behaviour in primary schools concerned with hitting their Key Stage 2 targets. Burgess et al (2005) examine whether the %5A*-C target indicator influences the distribution of student achievement as a result of the incentive to focus effort and resources on borderline pupils in order to boost the school's ranking.[1] Using pupil-level data for a cohort of all students in state maintained secondary schools in England, they find that both the educational gains and the exam performance of very low ability pupils are adversely affected. The effects found in this quantitative study are corroborated by the findings from a study that carried out in-depth interviews with secondary school headteachers (Wilson et al 2006). One question asked was: 'Have you ever targeted resources at particular areas in order to try and improve the school's position in the league tables? If so, how?' Out of a sample of 21, ten respondents discussed the strategies they had in place to target underachievement across the full ability distribution, regardless of their league table position. Eight headteachers

---

1   %5A*-C is the percentage of a school's pupils who obtain at least five 'good' GCSEs, i.e. at grade C or above. This has been the headline indicator in secondary school league tables in England since they were first published in 1992.

stated that they did currently target resources at C/D borderline pupils in order to improve their league table position, a further two said they had done so in the past, and one said s/he had tried, but failed, to avoid doing this. Conversely two said they simply didn't have any spare resources to target in such a way, and six stated that they deliberately didn't follow such a strategy. As one respondent said:

> "By choosing five or more A to Cs and making such a big issue out of it, it's no surprise that most schools put a huge amount of energy and resources into those students who are on the C/D borderline."

And when asked whether such efforts were successful the headteacher of a co-educational comprehensive school replied:

> "It is successful; as an educationalist I'm not entirely happy with it because I'm still thinking we're putting all these extra resources in these youngsters to keep A to C figures up, but maybe there's a youngster who's going to get an E which is really good for them and who might end up getting an F because we're not spending as much time with them. … The bright kids still prosper … I don't think they miss out at all. But I think the lower ability ones potentially do."

In summary, student outcomes as measured by raw test scores have improved since the introduction of performance management in both the UK and the US. Whether this represents an improvement in 'the quality of schooling or how much is learned each year' (Hanushek 2002: 2056) is more difficult to judge, as is the extent to which the observed gains can be attributed to specific forms of PI rather than other, often concurrent, reforms.

Health

Despite more than a decade of experience with public disclosure of performance data in the US health care system, there has been little rigorous evaluation of its impact on outcomes, partly due to the complexities of their measurement (Marshall et al 2000). The most reliable US evidence comes from studies from the New York Cardiac Surgery Reporting System. Hannan et al (1994) found that mortality declined significantly following publication of data on mortality rates (and did not

find evidence of a cream skimming response to public disclosure of that data). Dranove et al (2003) use the same data to examine the impact of report cards on a range of outcomes, and conclude that overall, report cards reduce patient welfare, but could have more positive impacts in the long term.

Marshall et al (2003) discuss some observational studies on the effect of PIs on processes related to health outcomes, such as vaccination and mammography. Bost (2001), for example, shows that US health plans which publicly reported their data showed greater relative improvement on some measures over a three-year period compared to those plans which measured but did not publish performance data. Longo et al (1997) find improvements in obstetrics care processes for hospitals identified as being low quality, and that this effect was stronger in more competitive markets. Improvements in processes related to patient experience (as well as reductions in within-hospital waiting times) were identified as one result of a benchmarking exercise across emergency care departments in 12 Swiss hospitals (Schwappach et al 2003).

Recent UK evidence on the impact of targets, coupled with strong penalties for managerial failure, on hospital waiting times (a measurable, and widely used, indicator of health care quality) is provided by Propper et al (2008a; 2008b). They use data for Scottish and English hospitals, exploiting the 'natural experiment' of the common policy environment before devolution and the policy divergence post devolution across the two countries. From 2000 there was a regime of 'targets and terror' for waiting times for inpatient care in English hospitals but not in Scotland. They find that this system of targets worked: waiting times in England fell by 13 days at the mean, with the largest falls for the longest waits. Moreover, they find no evidence of this being at the expense of other, less well-monitored, aspects of patient care, although there is some evidence of waiting list manipulation. Alvarez-Rosete et al (2005), Bevan and Hood (2006b) and Hauck and Street (2007) also compare waiting times across the devolved administrations of the UK and find that English waiting time targets had a positive impact.

Targets also appear to have had a positive impact on output in the primary care sector (Hood et al 2009). The Quality and Outcomes Framework (QOF) introduced for family doctors (GPs) in the UK in 2004 involved payments linked to specific measures of GP activity, and GPs on average greatly exceeded the targets set for the treatments that were measured and incentivised, which in fact led to a major overspend on the relevant NHS budget in 2005 (National Audit Office 2008). There

is, however, also evidence of output distortions resulting from this performance regime: non-incentivised treatments and holistic patient care received lower priority as GPs focused their efforts on treatments that triggered a payment (Checkland et al 2008).

Before discussing gaming responses in more detail, it is worth briefly considering the evidence on the degrees of response by different stakeholders to public disclosure of information in the health care sector. Marshall et al (2000) identify possible responses to public disclosure from consumers, hospitals and provider organisations, and purchasers. While consumers claim to want such information they do not make great use of the data (Propper and Wilson (2006) and references therein; see also Wilson (2010) for more on this). The evidence suggests that physicians are more aware of report cards than consumers but make little use of them (Marshall et al 2003). There are examples from the US of physicians responding by demanding that their managers' performance also be judged using report cards (Kaplan et al 2000). In the UK some professional networks (e.g. cardiothoracic surgeons) are showing some support for benchmarking on the basis of published data, although a recent study found some evidence that individual doctors perceived such disclosure as a threat to their autonomy (Exworthy et al 2010).

Provider organisations and hospitals are the most responsive agents to these data. Marshall et al (2000) indicate some positive responses, in that report cards induced better outcomes, particularly in competitive markets (see also Marshall et al 2003 and references therein). Other responses were less positive and included criticisms of the data, particularly among those hospitals identified as poor quality (Mason and Street 2005). Purchasers, in contrast, are more interested in costs or gross indicators of quality (Marshall et al 2003).

There are many examples of different types of gaming response to health care PIs. Green and Wintfeld (1995) argue that 41 percent of New York State's reduction in risk-adjusted mortality could be accounted for by data gaming. For example, Burack et al (1999) surveyed 104 cardiac surgeons in New York and found that high risk CABG (coronary artery bypass graft) patients were more likely to be denied treatment compared to similar high risk patients with aortic dissection, because the former procedure was subject to performance monitoring which created the incentive for surgeons to avoid treating high risk individuals.

For the UK, Bevan and Hood (2006a) report evidence on five types of output-distorting gaming response resulting from hospital A&E waiting time targets. These

include drafting in extra staff and cancelling operations scheduled for the period over which performance was measured; requiring patients to wait in ambulances until they were guaranteed to be seen within the four hour target; and removing the wheels from trolleys to turn them into 'beds' to satisfy the target that patients must be admitted to a hospital bed within 12 hours of emergency admission (see Bevan and Hood (2006a) for more examples and references). Bevan and Hamblin (2009) provide evidence of gaming by ambulance trusts: about a third of trusts show sharp discontinuities in the frequency distribution of ambulance response times for life-threatening emergency calls at eight minutes – the target level – strongly suggesting that some response times had been 'adjusted' to meet that target. Hood (2006) reports further evidence of ratchet effects, threshold effects and output distortion from interviews with 89 UK upper-level central government officials involved in managing the target system.

In summary, it is again difficult to tell which aspect(s) of measurement error, motivation and strategic response lead to the outcomes observed in the health care sector, which makes isolating and evaluating the impact of any one element of metrics-based accountability extremely challenging.

Discussion

So what have we learnt from the use of PIs in public service delivery to date? Providers do respond to the public disclosure of performance information when it is part of an incentive structure comprising financial and/or non-financial rewards and sanctions. There is some evidence that targets and such 'carrots and sticks' work, particularly if the desired outcome is focussed and measurable, as in the case of hospital waiting times. The two assumptions underlying such governance structures don't hold for public service delivery, however: measurement error is an inherent problem, as is the resultant potential for undesired as well as desired responses, and the evidence bears this out. But the alternatives to PIs in the management of public services aren't problem-free either, and the pressure to develop management-by-numbers systems is not likely to disappear. Here therefore are some policy issues that are relevant for the short to medium term.

The increasing use of PIs over the last ten-plus years, in the UK at least, has been against a background of 'good times', with public service performance being measured and monitored in the context of rising budgets. Now in the 2010s the fiscal landscape is somewhat different, which raises questions around what should be measured and how many resources the state should devote to measurement

(Hood et al 2009). Greater fiscal pressure is likely to mean more focus on input reduction and productivity targets, and indeed a recent policy document does reflect this, with an increased emphasis on 'value for money'; on taking account of the cost of achieving improved outcomes (HM Government 2009). How this may be achieved, given the inherent complexities involved in measuring both costs and benefits is less clear. It may also be politically prudent to have less emphasis on the use of 'user satisfaction' as an indicator of public service quality, given the difficulties of having to provide more for less, but there is no evidence of this trend so far. Costly measurement systems may come under greater pressure to justify themselves when the resources involved could be put to other 'front line' uses. We may, therefore, expect pressure on these forms of regulation to increase, and/or for more substantial evidence regarding their relative costs and benefits to be required through piloting of such systems before their more general introduction (Propper and Wilson 2003). Or the extent to which ranking exercises and similar need to be conducted by the state may be questioned: rankings of hospitals, universities and cities are carried out by independent organisations in some countries.

There is now a substantial body of evidence regarding the strengths and weaknesses of the different types of PI, and a range of options have been proposed for modifying them (Bevan and Hood 2006a). There may additionally be scope for the development of targets based on alternative, independent and therefore 'non-corruptible' information sources to reduce the potential for gaming (Propper and Wilson 2003). Legal review may also be an alternative form of incentive to drive performance improvement (Landman et al 2009). While the question of when 'good' data is good enough remains an open one, it is clear that better data is a necessary – but not sufficient – condition for better managing by numbers. Two points are worth emphasising here. First, there is more scope for learning from comparison, across the devolved administrations of the UK for example. But for such comparison to be effective there needs to be commensurability of the indicators used, across organisations; across public service domains; across jurisdictions. One theme that has emerged from the ESRC Public Services Programme (www.publicservices. ac.uk) is that there are still substantial barriers to such commensurability, which will limit the effectiveness of the use of comparative data (Wilson 2010b). There may be scope for the development of 'data clubs' to aid comparison across organisations and/or jurisdictions, which would involve mutually agreed arrangements for data-sharing, common conventions or methods of translation, and shared terms of use.

The long term payoffs of such an arrangement could be substantial (Hood et al 2009). My second point here is summed up by a footnote in small print to a 4-page spreadsheet of data in one of the annexes to the recent 'Smarter Government' policy document (HM Government 2009): "Exercise care and judgement in comparing data. See 'Benchmarking the Back Office: Central Government' for details and commentary". Data can be misrepresented, misused and misunderstood. Simply publishing it may not lead to the desired outcomes, particularly given the range of uses to which it is put and the range of stakeholders who have potentially different interests in the outcomes. Learning and education in what the data means is required, both for professionals and for citizens or 'consumers'.

More generally, more understanding is needed of how best to link measurement with management in public services: of when to use which PI(s), in isolation or in combination with alternative governance structures. PIs do not necessarily need to be coupled with targets, rankings or other metrics-based accountability mechanisms, but could be used as part of a system of 'intelligence' (Hood et al 2009). Such a system would have no automatic and therefore no predictable effects for providers and could not so easily be gamed. It could, however, still be 'high stakes': the metrics information could be used as the basis for intervention at various levels, and could be attached to rewards and/or sanctions, as well as to learn about improvement and evaluate policy. For example, an absolute categorisation of schools on the basis of their contextual value added 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 (Wilson and Piebalga 2008). This seems a more sensible use of this performance data than the somewhat spurious ranking of individual schools that is the current focus. An intelligence-based approach, perhaps combining metrics with site visits, may also be better able to deal with the challenges created by the recent shift towards area based or cross-cutting outcomes, i.e. those that involve multi-agency working which increases the difficulties of attributing risk, effort, and impact in the delivery of such outcomes.

Public agencies exist partly because there are conflicting goals among the stakeholders they represent, so any one set of PIs will not solve the problems of governance in such bureaucracies. A key challenge now is to use PIs more intelligently in the context of the broader performance management regimes of which they form a part.

References

Allen, R and Burgess, S (2010) *The future of competition and accountability in education,* mimeo.

Alvarez-Rosete, A, Bevan, G, Mays, N and Dixon, J (2005) Effects of diverging policy across the NHS, *British Medical Journal*, 331: 946-950.

Bevan, G and Hood, C (2006a) What's measured is what matters: targets and gaming in the English public health care system, *Public Administration*, 84(3): 517-538.

Bevan, G and Hood, C (2006b) Have targets improved performance in the English NHS? *British Medical Journal*, 332: 419-422.

Bevan, G and Hamblin, R (2009) Hitting and missing targets by ambulance services for emergency calls: effects of different systems of performance measurement within the UK, Journal of the Royal Statistical Society Series A, 172(1): 161-190.

Bost, J (2001) Managed care organizations publicly reporting three years of HEDIS data, *Managed Care Interface*, 14: 50-54.

Bradley, S, Crouchley, R, Millington, J and Taylor, J 9(2000) Testing for quasi-market forces in secondary education, *Oxford Bulletin of Economics and Statistics*, 62(3): 357-390.

Burack, JH, P Impellizzeri, P Homel and JN Cunningham Jr (1999) Public reporting of surgical mortality: a survey of New York State cardiothoracic surgeons, *Annals of Thoracic Surgery*, 68: 1195-1200.

Burgess, S, Propper, C and Wilson, D (2002) *Does performance monitoring work? A review of evidence from the UK public sector, excluding health care*, CMPO Discussion Paper 02/49, Bristol: CMPO, University of Bristol.

Burgess, S and Ratto, M (2003) The role of incentives in the public sector: issues and evidence, *Oxford Review of Economic Policy,* 19(2): 285-300.

Burgess, S, Propper, C, Slater, H and 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, Bristol: CMPO, University of Bristol.

Burgess, S, Propper, C, Ratto, M, von Hinke Kessler Scholder, S and Tominey, E (2009) *Smarter task assignment or greater effort: the impact of incentives on team performance*, CMPO Discussion Paper 09/215, Bristol: CMPO, University of Bristol.

Checkland, K, Harrison, S, McDonald, R, Grant, S, Campbell, S and Guthrie, B (2008) Biomedicine, holism and general medical practice: responses to the 2004 General Practitioner contract, *Sociology of Health and Illness*, 30(5): 788-803.

Clotfelter, C and Ladd, HF (1996) Recognizing and rewarding success in public schools, in: Ladd, HF (ed) *Holding Schools Accountable,* Washington DC: Brookings Institution, 23-63.

Crewson, PE (1997) Public service motivation: building empirical evidence of incidence and effect, *Journal of Public Administration Research and Theory,* 7(4): 499-518.

Cullen, JB and Reback, R (2002) *Tinkering towards accolades: school gaming under a performance accountability system*, University of Michigan, March.

Deci, EL (1971) Effects of externally mediated rewards on intrinsic motivation, *Journal of Personality and Social Psychology*, 18(2): 105-115.

Dee, T and Jacob, B (2009) *The impact of No Child Left Behind on student achievement,* NBER Working Paper 15531, Cambridge MA: NBER.

Deere, D and Strayer, W (2001) *Putting schools to the test: school accountability, incentives and behaviour*, Department of Economics, Texas A&M University, March.

Dixit, A (2002) Incentives and organizations in the public sector: an interpretive review, *Journal of Human Resources*, 37(4): 696-727.

Dranove, D, Kessler, D, McClellan, M and Satterhwaite, M (2003) Is more information better? The effects of 'report cards' on health care providers, *Journal of Political Economy,* 111(3): 555-588.

Exworthy, M, Gabe, J, Rees Jones, I and Smith, G (2010) *Opening up? Disclosure of surgeons' clinical performance and the impact upon their professional autonomy*, paper prepared for the IRSPM Conference, Berne, April 2010.

Figlio, DN and Getzler, LS (2002) *Accountability, ability and disability: gaming the system?* NBER Working Paper 9307, Cambridge MA: NBER.

Figlio, DN and Winicki, J (2005) Food for thought: the effects of school accountability plans on school nutrition, *Journal of Public Economics*, 89(2-3): 381-394.

Fitz-Gibbon, CT and Tymms, P (200) Technical and ethical issues in indicator systems: doing things right and doing wrong things, *Educational Policy Analysis Archives*, 10(6), available at http://epaa.asu.edu/epaa/v10n6.

Frey, B (2000) Motivation and human behaviour, in: Taylor-Gooby, P (ed) *Risk, Trust and Welfare,* Basingstoke: MacMillan: 31-50.

Gewirtz, S, Ball, SJ and Bowe, R (1995) *Markets, Choice and Equity,* Milton Keynes: Open University Press.

Green, J and N Wintfeld (1995) Report cards on cardiac surgeons: assessing New York State's approach, *New England Journal of Medicine,* 332: 1229-1232.

Gregg, P, Grout, P, Ratcliffe, A, Smith, S and Windmeijer, F (2008) *How important is pro-social behaviour in the delivery of public services?* CMPO Discussion Paper 08/197, Bristol: CMPO, University of Bristol.

Hannan, EL, Kilburn, H, Racz, M, Shields, E and Chassin, MR (1994) Improving the outcomes of coronary artery bypass surgery in New York State, *Journal of the American Medical Association*, 271: 761-766.

Hanushek, EA and Raymond, ME (2005) Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, 24(2): 297-327.

Hauck, K and Street, A (2007) Do targets matter? A comparison of English and Welsh National Health priorities, *Health Economics*, 16(3): 275-290.

Heckman, J, Heinrich, C and Smith, J (1997) Assessing the performance of performance standards in public bureaucracies, *American Economic Review, Papers and Proceedings,* 87: 389-395.

Heckman, J, Heinrich, C and Smith, J (2002) The performance of performance standards, *Journal of Human Resources,* 37(4): 778-811.

Heinrich, C (2003) Measuring public sector performance and effectiveness, in Peters, BG and Pierre, J (eds) *Handbook of Public Administration*, Thousand Oaks, CA: Sage.

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.

Hood, C, Dixon, R and Wilson, D (2009) *'Managing by numbers': the way to make public services better?* Oxford: ESRC Public Services Programme, University of Oxford.

HM Government (2009) *Putting the frontline first: smarter government*, Cm 7753, London: HMSO.

Jacob, BA (2002) *Acountability, incentives and behaviour: the impact of high-stakes testing in the Chicago public schools*, NBER Working Paper 8968, Cambridge MA: NBER.

Jacobs, R and Goddard, M (2007) How do performance indicators add up? An examination of composite indicators in public services, *Public Money and Management*, 27(April): 103-110.

Kane, TJ and Staiger, DO (2002) The promise and pitfalls of using imprecise school accountability measures, *Journal of Economic Perspectives,* 16(4): 91-114.

Kaplan, J et al (2000) Managed care reporting cards: evaluating those who evaluate physicians, *Managed Care Interface*, 13: 88-94.

Ladd, HF (1999) The Dallas school accountability and incentive program: an evaluation of its impacts on student outcomes, *Economics of Education Review*, 18(1): 1-16.

Ladd, HF (2002) *Market-based Reforms in Urban Education*, Washington DC: Economic Policy Institute.

Landman, T, Calvo, K, Platt, L and Sunkin, M (2009) Litigation and the quality of local government service delivery, PSP Discussion Paper 0808, Oxford: ESRC Public Services Programme.

Le Grand, J (2003) *Motivation, Agency and Public Policy*, Oxford: Oxford University Press.

Longo, DR, G Land, W Schramm, J Fraas, B Hoskins and V Howell (1997) Consumer reports in health care: do they make a difference in patient care? *Journal of the American Medical Association*, 278(19): 1579-1584.

Marshall, M, Shekelle, P, Brook, R and Leatherman, S (2000) *Dying to know: public release of information about quality of health care,* London: Nuffield Trust.

Marshall, M, Shekelle, P, Davies, H and Smith, P (2003) Public reporting on quality in the United States and the United Kingdom, *Health Affairs,* 22(3): 134-148.

Mason, A and A Street (2005) *Publishing hospital outcomes data in the UK*, London, Office of Health Economics.

McLean, I, Haubrich, D and Gutiérrez-Romero, R (2007) The perils and pitfalls of performance measurement: the CPA regime for local authorities in England, *Public Money and Management*, 27(2): 111-118.

Meyer, RH (1997) Value-added indicators of school performance: a primer, *Economics of Education Review*, 16(3): 283-301.

National Audit Office (2008) *NHS Pay Modernisation: New Contracts for General Practice Services in England*, TSO.

Neal, D and Schanzenbach, D (2007) *Left behind by design: proficiency counts and test-based accountability*, NBER Working Paper 13293, Cambridge MA: NBER.

Propper, C and Wilson, D (2003) The use and usefulness of performance measures in the public sector, *Oxford Review of Economic Policy*, 19(2): 250-267.

Propper, C and Wilson D (2006) The use of performance measures in health care systems, in: Jones, AM (ed) *The Elgar Companion to Health Economics*, Cheltenham: Edward Elgar, 326-334.

Propper, C, Sutton, M, Whitnall, C and Windmeijer, F (2008a), Did 'targets and terror' reduce waiting times in England for hospital care?, *The B.E. Journal of Economics Analysis and Policy,* 8(2)(Contributions), Article 5. Available at: http://www.bepress.com/bejeap/vol8/iss2/art5 (accessed 10 January 2010).

Propper, C, Sutton, M, Whitnall, C and Windmeijer, F (2008b) *Incentives and targets in hospital care: evidence from a natural experiment,* CMPO Discussion Paper 08/205, Bristol: CMPO, University of Bristol.

Reback, R (2008) Teaching to the ranking: school accountability and the distribution of student achievement, *Journal of Public Economics,* 92(5-6): 1394-1415.

Rowan, K, Harrison, D, Brady, A, and Black, N (2004) Hospitals' star ratings and clinical outcomes: ecological study', *British Medical Journal*, 328: 924-925.

Schwappach, D, A Blaudszun, D Conen, H Ebner, K Eichler and M Hochreutener (2003) 'Emerge': benchmarking of clinical performance and patients' experiences with emergency care in Switzerland, *International Journal for Quality in Health Care,* 15(6): 473-485.

Smith, PC (1995) On the unintended consequences of publishing performance data in the public sector, *International Journal of Public Administration*, 18(2/3): 277-310.

Stevens, P, Stokes, L and O'Mahony, M (2006) Metrics, targets and performance, *National Institute Economic Review*, 197: 80-92.

Tymms, P (2004) Are standards rising in English primary schools?, *British Educational Research Journal*, 30(4): 477-494.

West, A and Pennell, H (2000) Publishing school examination results in England: incentives and consequences, *Educational Studies*, 26(4): 423-436.

Whitty, G, Power, S and Halpin, D (1998) *Devolution and Choice in Education: the School, the State*

and the Market, Milton Keynes: Open University Press.

Wiggins, A and Tymms, P (2002) Dysfunctional effects of league tables: a comparison between English and Scottish primary schools, *Public Money and Management*, 22(1): 43-48.

Wilson, D (2004) Which ranking? The impact of a 'value added' measure of secondary school performance, *Public Money and Management,* 24(1): 37-45.

Wilson, D (2009) Exit, voice and quality in the English education sector, *Social Policy and Administration*, 43(6): 571-584.

Wilson, D (2010a) *Fostering citizen control,* mimeo.

Wilson, D (2010b) *Comparative analysis in public management*, paper prepared for the IRSPM Conference, Berne, April 2010.

Wilson, D and Piebalga, A (2008) Performance measures, ranking and parental choice: an analysis of the English school league tables, *International Public Management Journal,* 11(3): 1-23.

Wilson, D, Croxson, B and Atkinson, A (2006) 'What gets measured gets done': headteachers' responses to the English secondary school performance management system, *Policy Studies*, 27(2): 153-171.

2
Fostering citizen control

Abstract

Since the 1980s user-focused mechanisms such as choice, voice and personalisation have been introduced by governments as part of successive public service reforms, both in the UK and elsewhere. Such initiatives generally have the aim of improving public service quality through consumer pressure, as well as empowering users, making public services more responsive to their needs. As one policy document states "Bottom-up pressure through choice and voice can … give everyone, including the disadvantaged, better quality services" (PMSU 2006, page 10). This view that choice and voice can work together to improve quality for all users can be traced back to Hirschman (1970); however in a previous paper I argue that it may not hold in the current system of school choice as it operates in England (Wilson 2009).

In this paper I discuss the various goals of user-based mechanisms and show how the general terms of choice and voice are manifested in different ways depending on the specific welfare context. I briefly review the evidence on the use of choice and voice based mechanisms across different public service domains, which again highlights the importance of context on outcomes. I argue that the original Hirschman framework provides one way in which we can think about these differences within the same coherent framework and use education as a case study to illustrate this. I argue that user-based mechanisms cannot be judged without considering the broader performance management regimes of which they often form a part. I highlight the importance of the way 'quality' is measured and the incentives created by the type of performance information available to users. The nature of the public service 'product' and the constraints faced by users in realising their choices also impact on the outcomes of such schemes.

Introduction

Since the 1980s user-based mechanisms such as choice, voice and personalisation have been introduced by governments as part of successive public service reforms, both in the UK and elsewhere. Such initiatives generally have the aim of improving public service quality through consumer pressure, as well as empowering users, making public services more responsive to their needs. Choice and voice are considered to be complementary: "Choice and voice should complement each other ... Bottom-up pressure through choice and voice can ... give everyone, including the disadvantaged, better quality services" (PMSU 2006: 10). This discourse about choice and voice working together to improve quality can be traced back to Hirschman (1970), who argues that exit (choice) and voice are two consumer responses to deterioration in the quality of a firm's product or service. These provide signals to the firm, which responds by improving quality, thereby creating a self-correcting mechanism via which quality standards are maintained. Hirschman argues that different combinations of exit and/or voice are suitable in different settings, depending on which signal(s) the firm is most responsive to.

Bottom-up pressure from users is just one of four elements of the UK government framework for improving quality in public service delivery. The model of public service reform also incorporates top-down performance management; competition/contestability; and increasing the capability and capacity of public servants. I discuss top-down performance management in detail in a companion paper (Wilson 2010). It is recognised that this general model needs to be tailored to each service; in particular "The appropriate mix of top-down pressure, competition and bottom-up choice and voice will therefore vary from case to case" (PMSU 2006: 11).

In this paper I discuss the various goals of user-based mechanisms and show how the general terms of 'choice' and 'voice' are manifested in different ways depending on the specific welfare context. I briefly review the evidence on their use across different public service domains, which again highlights the importance of context on outcomes. I argue that the original Hirschman framework provides one way in which we can think about these differences within the same coherent framework and use education as a case study to illustrate this. I argue that user-based mechanisms cannot be judged without considering the broader performance management regimes of which they often form a part. I highlight the importance of the way 'quality' is measured and the incentives created by the type of performance

information available to users. The nature of the public service 'product' and the constraints faced by users in realising their choices also impact on the outcomes of such schemes.

Goals of user-based mechanisms

User empowerment through choice and voice may be desirable in and of itself. As Appleby et al (2003), quoted in Greve (2009a: 543) state 'who could argue against the desirability of allowing patients more say in decisions concerning them'. This may be particularly relevant in an age of a more 'assertive citizen' (Social Market Foundation 2009) who welcomes personal choice and a move away from one-size-fits-all provision. A recent survey found that the majority of people in Britain think parents should have choice over which schools their child attends, and that patients should have choice over which hospital provides their treatment (Curtice and Heath 2009). But this position is not undisputed. Users may find it intimidating and stressful to make choices, particularly given the (at least perceived) risk of making the 'wrong' choice; and voice may be time-consuming and frustrating (Greve 2009b; Greener 2007a; Schwartz 2004).

It is also hard to imagine that users will continue to be satisfied with having some form of control in their public services if there are no noticeable improvements in their experiences of them. Indeed, on the basis of the findings from his survey, John Curtice concludes that choice is not valued highly enough by people to provide sufficient justification for its use if it does not also create 'extrinsic benefits', i.e. generate improvements in service delivery.[2] A recent report from the 2020 Public Services Trust reaches a similar conclusion (2020 Public Services Trust 2010). This relates to the distinction made by Dowding and John (2009a) between intrinsic value – choice being desired for itself; and instrumental value – if it results in welfare gains in the form of productive and/or allocative efficiency. They too argue that attention should be concentrated on the potential welfare gains of choice rather than any supposed intrinsic benefits.

The use of choice and, increasingly, voice as means to achieve such welfare gains is now well established in the UK and elsewhere (Burgess et al 2007; Propper et al 2006; Simmons et al 2009). Le Grand (2007) argues that such mechanisms create a system with incentives for reform embedded within it. Choice, coupled

---

2 Presentation at the ESRC Public Services Conference, 11 December 2009. Details available from http://www. publicservices.ac.uk, accessed 1.2.10.

with a system of funding in which resources follow the user, applies competitive pressure to providers who therefore have the incentive to improve their service in order to attract or retain clients. Such improvements may include increased responsiveness to the needs of individual clients, which is often discussed as part of the 'personalisation' agenda. Another goal of user-based mechanisms has been to improve equality of access for those users who historically have found it harder to access good quality services. Le Grand (2007) goes further and argues that choice is better than voice at achieving this. So, for example, school choice can, in theory, break the link between house prices and access to 'good' schools by reducing the importance of geographical proximity as an allocation criterion. In practice this has not been so successful (Burgess et al 2009a; b).

Beneath this level of generality, however, there are both differences in aims and differences in emphasis for those aims across different public service domains. Perri 6 (2003) reviews choice in the UK across nine public service domains over a period of up to 15 years and identifies nine goals of individual consumer choice, including satisfaction, convenience, efficiency and responsiveness.[3] Greener and Powell (2008) provide an account of how choice policies in housing, education and health have evolved in the UK since the 1940s and show that policy documents in the three domains have emphasised different aspects of choice. In housing the emphasis has been on choice and responsibility; in education on choice and diversity; in health on choice and responsiveness. Greener and Powell argue, however, that what is less clear from the policy documents is how any of these different goals link to improved standards in each of the domains. Differences in goals are related to differences in the ways user-based mechanisms are both designed and realised in different institutional settings, and I look at this in the next section.

What is 'choice' and 'voice'?

Hirschman (1970) argues that a process of decline in the quality of a firm's output (for whatever reason) activates certain consumer responses which in turn act as endogenous forces of recovery, thereby reversing the initial decline in quality. This is a self-correcting mechanism, whereby the very process of decline activates certain counterforces and hence generates its own cure – the 'embedded incentives' in Le Grand's (2007) terminology. He distinguishes two contrasting consumer responses

3 The full list is: at the consumer level – outcomes, acceptability, satisfaction, user convenience. At the service level – experiment, responsiveness, efficiency, clientele, presentation.

– exit and voice. Exit is "the sort of mechanism economics thrives on" (Hirschman 1970:15). It is neat, impersonal and indirect: subsequent recovery by the firm comes via the market. Voice, by contrast, is more "messy", more personal and more direct, and for Hirschman can cover anything from personal complaint to collective action. So exit belongs in the economic realm, with consumer decisions in the marketplace as levers of change; while voice is positioned in the political, democratic realm with individuals as citizens making their feelings known (Dowding and John 2007; Greener 2007b).

Numerous variations and modifications of Hirschman's original exit-voice distinction have been put forward (I discuss the place of loyalty in this framework below). For example, Dowding and John (2008; 2009b) propose a 'three exit, three voice and loyalty' framework. They distinguish three ways an individual can exit from a public service provider. She can move locality; switch between public providers without moving; or switch from public to private provider. There are also three ways an individual can exert voice. Individually, through complaint; or collectively, through voting, or by joining some form of collective action such as campaigning or petitioning. Their analysis highlights the complex dynamics between the different forms of choice and voice, as well as between each potential response and an individual's (dis)satisfaction with public services. 'Bridging notions' that span exit and voice have also been proposed, including co-production, which may be particularly relevant to a long-term relationship between user and provider (Greener 2007b).

In practice, therefore, despite the simple appeal of 'choice' and 'voice', the terms are actually used in many different ways and can refer to quite different institutional arrangements, which in turn impact on the outcomes of any one user-based mechanism. Three questions serve to frame the basic elements of, for example, a choice-based system. First, who chooses? In education it is the consumer herself, the pupil, or the parents acting on her behalf. In the health care context the choice is more usually delegated to a specialised agent (gatekeeper, intermediary, payer) acting on the consumer's behalf. In practice, in health care, patient choice is often somewhat limited, with payer choice dominant, although the distinction is sometimes blurred (Propper et al 2006).

Second, what do they choose? In state-funded systems, price is generally not the determining factor and quality is important. Parents are likely to care about the quality of education, broadly defined, which may include factors such as a school's

previous results, its composition, ethos, facilities, location (Wilson 2009; Burgess et al 2009a; b; Allen and Burgess 2010). In health care, the patients are often fully insured against the price and so choose on the basis of quality, while the body responsible for buying care will be interested in both.

Finally, what are the constraints on the process of choice? For parents the most obvious constraint is that their preferred school is full. Transport time and costs may be a constraint on choosing other schools, and house prices may be a constraint on moving. In health care, travel costs and other costs of using care may constrain choice. Consumers may feel 'locked-in' to local government services, with little realistic option for exit. In all public service contexts the lack of accurate information on the quality of the often complex service being provided may limit the extent to which consumer choice, or voice, creates the incentive to improve that quality, and I discuss this further below.

Overall, as Greve (2009b) states, both the conditions and context for choice have an impact on what type of choice is made, and how. Blank (2009) distinguishes three inter-related elements of that context: the public institutions that give markets in each domain distinctive features; the individual cognitive and material preconditions to deal successfully with choices from alternative providers; and the features of goods/services that are the subject of choice. All these contribute to differences in the outcomes of user-based mechanisms across different public service domains.

Evidence on user-based mechanisms

There is now a substantial body of evidence on choice in public services (Burgess et al 2007; Propper et al 2006; 6, 2003) but much less on voice, or on how choice and voice may work together, although the work of Dowding and John (2008; 2009) discussed above represents a recent contribution. My aim here is not to provide a full review the evidence that has been summarised elsewhere; rather I want to focus on what the literature tells us about if and when people want choice or voice; how they respond to that opportunity and how providers may respond to the signal(s) provided.

Whether we think of user-based mechanisms solely as means of empowering users or, as is argued above, as a means by which the quality of public services are improved, a fairly obvious consideration is whether people want choice, and, if so, (how) do they respond? Curtice and Heath (2009) and 2020 Public Services Trust

(2010) both find that people want choice, but that it is not necessarily a priority. They find less support for diversity of provision, with only a minority of respondents in favour of private companies running either (state funded) hospitals or schools. The Social Market Foundation (2009) also report that, while users welcome personal choice and the subsequent potential for increased responsiveness in service provision, they have concerns regarding the possible inequity in service delivery (in the form of a 'postcode lottery' for example) that may result. With regard to voice, Greener (2007b) discusses experiments in health and education that aim to increase democracy, but notes that many commentators suggest such experiments will not change practice because of the lack of opportunity and time that local people have to become involved.

Propper et al (2006) review the evidence on whether patients respond to greater choice in health care (see also Dowding and John 2009a). They conclude that direct patient choice is limited in many systems, and that it may conflict with choice exercised by the agents who place contracts with hospitals on behalf of groups of patients. Patients in England have expressed willingness to travel to non-local hospitals and have done so when given assistance to exercise this choice. When such support is absent (or the wait at the local hospital is perhaps shorter or less uncertain), the evidence from European countries suggests there is relatively little take-up of such travel options. Individuals who are better informed and individuals whose illnesses are more severe may be more likely to travel and this may lead to greater differences across hospitals in patient severity.

This points to differences across individuals in their exercise of choice. There are also potential issues regarding differences in individuals' basis for choice. While the theory behind school choice, for example, rests on parents choosing at least partly on the basis of academic standards, which in turn creates the incentive for schools to improve those standards, it may be that parents (additionally) have what 6 has termed 'segregationist preferences' (6, 2003 and references therein; Allen and Burgess 2010), whereby individuals choose schools (or other public services such as housing) on the basis of trying to avoid certain types of other users. Recent evidence from Burgess et al (2009a; b) shows that parents choosing a primary school in England do value academic standards, but they also choose on the basis of proximity and a school's socio-economic composition, preferring schools with lower proportions of pupils from low income households.

More generally, if we consider individuals choosing on the basis of 'quality' of the public service, or the 'performance' of the public sector organisation, there are

issues around the complexity of (the measurement of) such outcomes, which are discussed in detail in the companion paper to the present one (Wilson 2010). Any measure of performance will only achieve a partial summary of the true quality of outcomes, and this has an effect on how both consumers and providers behave. The evidence from health care suggests that while consumers have access to information, it is often too complex for direct use by them or even by the buyers of health care acting on their behalf (Marshall 2002). There is also evidence that some patients need more guidance in processing and using such complex performance information, which has additional equity implications (Dowding and John 2009a). The problem of a potential lack of understanding by the public regarding what the published information means may be compounded by the publication of numerous different indicators of an organisation's performance. While 'value added' performance measures more accurately reflect a school's effectiveness, for example, the fact that parents are more familiar with the (easier to understand) raw outcome measures means that headteachers still have the incentive to focus on the latter, possibly at the detriment of true improvements in performance (Wilson et al 2006; Wilson 2010).

This then leads to the broader issue of whether and how providers respond to user-based mechanisms, which is important if such mechanisms are to lead to welfare gains. Do providers respond to choice and voice by improving the quality of their outcomes? And are there other, undesirable and/or unanticipated responses? The rather mixed evidence on choice has been reviewed elsewhere (Propper et al 2006; Burgess et al 2007; Propper and Wilson forthcoming; 6 2003; Allen and Burgess 2010). There is less evidence with regard to voice, although Greener (2007a) does comment that public services have historically been poor at dealing with voice mechanisms, especially in terms of holding professionals to account. A common theme that comes out of all the reviews is that it is extremely difficult to generalise about the impact of 'choice' or 'voice' on outcomes. The effects of user-based mechanisms depend to a large extent on system design and institutional context, both in terms of actual quality improvements achieved and other, less desired responses.

In summary, user empowerment is not a sufficient end in itself. User-based mechanisms need to lever improvement in public service quality; to create instrumental as well as intrinsic value (Dowding and John 2008). But 'choice' and 'voice' mean different things both in different welfare settings and, within the same

setting, to different individual users, and the available performance information impacts on both sides of any 'market'. The original analysis of Hirschman provides one way in which we can think about these differences within the same coherent framework.

Hirschman's exit, voice, loyalty and quality

Hirschman (1970) defines exit and voice as two possible user responses to a decline in the quality of a firm's product or service. His particular interest was how, and under what circumstances, exit and voice may combine to best rectify or reverse such a (relative or absolute) decline in quality:

> … "how a typical market mechanism and a typical non-market, political mechanism work side by side, possibly in harmony and mutual support, possibly also in such a fashion that one gets into the other's way and undercuts its effectiveness" (Hirschman 1970: 18).

To be effective, he argues, the signal used – exit or voice – should correspond with that to which the organisation is responsive, which in turn depends on the particular service/product and/or organisation being considered. I shall later argue that the responsiveness of the organisation depends at least in part on the incentives created by the broader performance management system within which the organisation operates (see also Paul 1992). Hirschman identifies a particular problem, however: that over-emphasis on less costly exit may reduce investment in (may 'atrophy') voice, even in circumstances when voice may be the most effective mechanism for improving quality. This is because those consumers who care most about quality – and who would be the most active agents of voice – are for that very reason those most likely to exit first when faced with a decline in that quality. If the firm is more responsive to voice, this will make the self-correction mechanism less effective at restoring levels of quality. Hirschman recognises a tension between exit and voice: consumers' willingness to develop and use the voice mechanism is reduced by exit, but the presence of an exit option increases the effectiveness of voice.

One way in which exit, particularly by quality-sensitive consumers, may be delayed is through what Hirschman (1970: 79) calls loyalty: "The importance of loyalty … is that it can neutralize within certain limits the tendency of the most quality-conscious customers or members to be the first to exit". There is a debate

in the literature regarding whether loyalty is psychological or behavioural (Dowding et al 2000; Dowding and John 2008; 2009b), but it can be understood as a barrier to exit which may be directly imposed or internally generated. Hirschman argues that staying within a declining organisation may in fact be rational if, by exiting, the quality of the organisation further declines *and* the consumer cares about its quality even after she's left.[4] This in turn implies she does not fully exit ("voice from within" compared to "voice from without"). Hirschman introduces the term "quality maker" to describe that situation where a consumer's exit causes quality to decline, and I return to this term below.

Hirschman identifies two scenarios with regard to quality. The majority of his analysis draws on the assumption that a change in quality is felt in the same direction by all consumers: individuals may be differentially sensitive to such a change, but all agree that it is either a decline or an improvement. I additionally use the term 'uni-dimensional' quality to describe this scenario: quality can improve or decline only along one dimension, and individuals all agree on the direction of change along that dimension. Hirschman more briefly considers the case when a change in quality is felt in different directions by different consumers: individuals may disagree on whether an increase in the level of a particular service is a good or a bad thing depending on their political affiliations, for example. I additionally use the term 'multi-dimensional' quality to describe this scenario, where the preferences of consumers may differ across alternative dimensions and hence they may disagree whether changes along any one dimension represent a decline or improvement in quality.[5] Crucially, Hirschman shows that the operation of exit and/or voice yields different outcomes depending on which concept of quality is relevant. I now examine each scenario in turn, using the case of education to illustrate how these different outcomes may be realised in a specific institutional setting.

A change in quality is felt in the same direction by all: uni-dimensional quality

The assumption that a change in quality is felt in the same direction by all consumers underlies much of Hirschman's analysis. In this case, consumers agree on whether a change is an improvement or a decline in quality, but they may be differentially sensitive to such change. If quality declines, exit and voice are complementary in

---

4 Francois (2000; 2001) analyses the effects of individuals placing a value on the quality of service provided even though they do not direct receive personal benefit. His focus is on how such 'care' impacts on employee motivation in the provision of public services.

5 This quality distinction parallels that between vertical and horizontal product differentiation in the economics literature (Gaynor 2006). See Wilson (2008) for more discussion on this point.

the sense that they both work to improve quality, as long as the signal used is that to which the organisation is responsive. There are spillovers or externalities between alert and inalert consumers: the latter benefit from the quality improvement brought about by the exit or voice of the former precisely because they all see it as an improvement. The presence of exit may reduce investment in voice, however: exit may atrophy voice. The most quality-sensitive, and therefore the potentially most vocal, are likely to be the first to exit, leaving behind less vocal consumers. This is a problem if the organisation is responsive to voice, as the signal it needs to improve quality will be weaker and the self-correction mechanism therefore less effective.

So if a quality change is felt in the same direction by all consumers, Hirschman argues that exit and voice are complementary and there are spillovers between alert and inalert consumers, but there is the potential for over-emphasis on exit even when the firm is more responsive to voice, which may prevent or delay recovery.

Application to the English education sector

Consider quality as school effectiveness, or value added, an improvement in which, I argue, is felt as such by all consumers (although not necessarily to the same degree). In this case, the Hirschman thesis suggests that exit and voice will indeed be complementary; that both these user-driven mechanisms will work together to provide signals to the schools to improve their effectiveness. The actions of the alert will cause spillovers for the inalert consumers, as all benefit from the agreed-upon improvement. This sounds very much like the "rising tide that raises all boats" scenario of Hoxby (2003).

But what about the problem of exit atrophying voice? Hirschman argues that the possibility of exit reduces investment in voice, and that the most vocal exit first which leaves less scope for effective voice. Whether or not this is a problem depends on how responsive schools are to the different signals, which in turn is determined by the incentives they face as part of the broader performance management regime within which choice and voice operate (Paul 1992). There is therefore a fundamental link between user-based mechanisms and more 'top-down' accountability regimes. The current system in England, in which school funding relies directly on pupil numbers, is one in which school incentives are based more on choice than on voice, so atrophy should not be a problem: the top-down performance management system creates the incentive for schools to respond to the choice signal. One implication of this, however, is

that if policy makers are attempting to introduce more options for voice as an additional user-driven mechanism alongside choice, they also need to create the incentive for schools to respond to that signal, and to respond by improving their effectiveness.

So *if* all key players are (only) interested in quality as value added or effectiveness, the Hirschman analysis concurs with current policy discourse in predicting that choice and voice should complement one another in improving quality for all. The design of the performance management system is central to the relative degree of responsiveness of schools to the two signals.

A change in quality is felt in different directions by different consumers: quality is 'multi-dimensional'

Hirschman also (but more briefly) considers the case when a change in quality is not appreciated as such by all consumers, i.e. when consumers have a differential appreciation of the same quality change. He gives the example of different political affiliations leading to differing views regarding changes in local government spending. We can similarly think in terms of quality being multi-dimensional: a change in quality along one dimension may be appreciated by some but not by others. In such a scenario, Hirschman argues, organisations have the possibility of changing quality in such a way as to please some while displeasing others. Which route will they take? To whom will they respond?

There are in fact no specific predictions arising from the Hirschman analysis regarding the answers to these questions. Rather, Hirschman discusses in general terms the 'quality path' of the organisation, and how this path depends on the organisation's responsiveness to exit and/or voice. For example, if it is more responsive to exit than voice, the organisation is more likely to correct deviations from normal quality that are 'obnoxious' to its exit-prone customers. This may not be seen as an improvement by its vocal customers. Alternatively, if the organisation is more responsive to voice, it may work to minimise discontent among its vocal customers by changing quality in ways that are not appreciated by those who are exit-prone. The quality path of the organisation can therefore be predicted in different contexts, or under different assumptions regarding the relative responsiveness of the organisation to the different signals. Crucially, if quality is multi-dimensional, exit and voice do not necessarily complement each other because exit-prone and more vocal consumers may view the same change in quality differently from one another.

Moreover, there will be no spillovers between alert and inalert consumers if they value different aspects of quality. Spillovers may be possible in a multi-dimensional quality setting, but only if alert and inalert users similarly value quality changes along the same dimension. They are no longer guaranteed.

Application to the English education sector

The notion of multi-dimensional quality seems in tune with the actual basis for choice of parents, who may have different preferences across different dimensions of school 'quality': test scores, composition, ethos, location. In this case the Hirschman analysis predicts that there is no guarantee that choice and voice will complement each other, nor that there will be spillovers between alert and inalert consumers if they value different aspects of quality. There is no longer any guarantee of that 'rising tide'.

Can we say anything about the likely outcome, about the 'quality path' schools have the incentive to take in each case? I argue that the notion of the pupil as a quality maker, and the fact that the headline performance indicators (PIs) of quality incorporate this, proves useful in predicting the outcome.

Raw output measures such as a primary school's average Key Stage 2 scores, or the proportion of a secondary school's students achieving five 'good' GCSEs, are still the key indicators in the English school league tables. School rankings in the league tables matter to all key players in the education system (Wilson et al 2006). Schools have the incentive to care about outcomes as measured by these summary indicators of raw test scores and therefore have the incentive to care about – to respond to signals from – pupils of high ability whose exit would reduce (or entry would enhance) quality as measured by such indicators, which directly relates to Hirschman's notion of consumers as quality makers. Specifically, they have the incentive to respond to parents of high ability children. Given the general positive association between income and attainment, these are going to be, broadly speaking, middle class parents. In the education context, the middle class parents are likely to be the most exit-prone and the most vocal (Le Grand 2007). Contrary to Hirschman, therefore, there may not be a conflict between responding to the (threat of) exit or responding to voice in the multi-dimensional quality setting. Instead, schools currently have the incentive to respond to *either* signal from the parents whose children will boost (measured) quality. This incentive comes from the design

of the performance management system and, in particular, the importance of performance measures that incorporate pupils as quality makers.

So in the education context choice and voice do complement each other, even though quality is multi-dimensional, *but* only for one type of consumer. And schools have the incentive to focus on the elements of quality preferred by that type of consumer. This provides one way of thinking about how to predict the resultant quality path followed by the school. For example, one aspect or dimension of quality which the school may subsequently have the incentive to change is composition. This provides potential links with the debates on (covert) selection by schools (Le Grand 2007). There has been recent evidence that some schools in England have been breaking admissions laws in ways which, according to Schools Minister Jim Knight, penalised poorer families.[6] Similarly, a government inquiry found that 17% of the 570 secondary schools checked in three local authorities were breaking the admissions rules, for example by asking parents banned questions about marital status and financial background, or by not giving due priority to children in care or with special needs.[7]

Discussion

While choice and voice are often associated with debates about user empowerment and control, this is not sufficient, at least in the long term. Such user-based mechanisms also need to create some form of welfare gain, towards the goal(s) of improving efficiency, equity of access, responsiveness. The evidence on the extent to which these goals are realised – on the extent to which public service performance is improved – is, however rather mixed. Indeed, a common theme that emerges from the literature is the extent to which the outcomes of any one user-based mechanism are situation specific, which suggests caution is required before translating a policy from one institutional context to another. This caution, according to Greener and Powell (2008), has not been exercised by policy makers.

Hirschman's original analysis provides one framework within which we can consider how choice and/or voice may – or may not – work to improve public service performance in different settings. It highlights several important features, both from the user and from

6   See http://news.bbc.co.uk/go/pr/fr/-/1/hi/education/7193052.stm (story published 17/01/2008; accessed 26/06/2008).
7   See http://news.bbc.co.uk/go/pr/fr/-/1/hi/education/7326347.stm (story published 03/04/2008; accessed 26/06/2008).

the provider perspective. What are people choosing? And are they likely to agree about the direction of change in performance? This may determine the extent to which the actions of the few benefit the many through positive spillovers. Or is 'quality' more multi-dimensional? In this case the notion of the 'quality path' helps in predicting likely provider responses in different welfare contexts. And how are provider incentives to respond to either choice or voice affected by the broader performance management regimes to which they are subjected? In particular, to what extent does the way 'quality' is (imperfectly) measured impact on both user and provider behaviour?

The answers to all these questions will be context-dependent. It is clear, however, from just using the very simple Hirschman framework that quite a specific set of conditions need to be in place for either choice or voice to work in the way that policy discourse suggests. This is true even without explicitly incorporating the potentially complex dynamic relationships between the two options in different welfare settings, or, crucially, without considering the constraints both that users face in producing the signals and providers face in responding to them.

So, for example, recent work by Burgess et al (2009 a; b) investigates parental preferences for primary schools. They find that families across the income distribution have broadly similar preferences across different dimensions of school quality, which may suggest that positive spillovers could be possible even in this multi-dimensional setting. But there are two, quite different, constraints that may prevent that 'rising tide'. First are the preferences themselves. While academic standards are valued, so is a student body comprising a low proportion of low income students. Such preferences for peer groups create pressure for socio-economic sorting. Because parents value academic standards, schools do have the incentive to increase their attainment (here, Key Stage 2) measure, but can do so by manipulating their intake peer group as well as raising effectiveness. Parents also value proximity. A key finding from this study, however, is that higher and lower income families face huge differences in the types of schools that are feasible for them to access. The big driver of differential access to better schools is the quality of schools nearby to where families live, coupled with the use of proximity – the distance from a family's home to the school – as a device to allocate places at over-subscribed schools. Such capacity constraints limit the extent to which individuals can realise their preferences for specific schools, and prevent choice from improving standards for all.

More generally, there are many other examples of constraints limiting the successful operation of user-based mechanisms in practice, including lack of time to become involved in voice-related activities; lack of understanding of complex performance information; limited opportunities for user choice in systems reliant on health or social care professionals; lack of flexibility with regard to both entry and exit into any specific welfare market. While there are potential solutions to all these constraints, they come at a cost, and whether it will be feasible, either economically or politically, for governments to bear such costs is far from certain. What this analysis shows, however, is that there are big differences between the rhetoric and the reality when it comes to giving control to citizens over their use of public services.

References

6, P (2003) Giving consumers of British public services more choice: what can be learned from recent history? *Journal of Social Policy*, 32(2): 239-270.

2020 Public Services Trust (2010) What do people want, need and expect from public services, http://www.2020publicservicestrust.org.

Allen, R and Burgess, S (2010) The future of competition and accountability in education, mimeo.

Appleby, J, Harrison, A and Devlin, N (2003) *What is the real cost of more patient choice?* London: King's Fund.

Blank, F (2009) When 'choice' and 'choice' are not the same: institutional frameworks of choice in the German welfare system, *Social Policy and Administration*, 43(6): 585-600.

Burgess, S, Greaves, E, Vignoles, A and Wilson, D (2009a) *What parents want: school preferences and school choice*, CMPO Discussion Paper 09/222, Bristol: CMPO.

Burgess, S, Greaves, E, Vignoles, A and Wilson, D (2009b) *Parental choice of primary school in England: what 'type' of school do parents choose?*, CMPO Discussion Paper 09/224, Bristol: CMPO.

Burgess, S, Propper, D and Wilson, D (2007) The impact of school choice in England: implications from the economic evidence, *Policy Studies,* 28(2): 129-143.

Curtice, J and Heath, O (2009) Do people want choice and diversity of provision in public services?, in: Park, A, Curtice, J, Thomson, K, Phillips, M and Clery, E (eds), *British Social Attitudes: The 25th Report*, London: Sage, 55-78.

Dowding, K and John, P (2008) The three exit, three voice and loyalty framework: a test with survey data on local services, *Political Studies,* 56(2): 288-311.

Dowding, K and John, P (2009a) The value of choice in public policy, *Public Administration,* 87(2): 219-233.

Dowding, K and John, P (2009b) *The three exit, three voice and loyalty framework: a test with panel survey data on urban services*, Paper presented at the Midwest Political Science Association Annual Conference, April 2009.

Greener, I (2007a) Choice or voice? Introduction to the themed section, *Social Policy and Society,* 7(2): 197-200.

Greener, I (2007b) Choice and voice: a review, *Social Policy and Society*, 7(2): 255-265.

Greener, I and Powell, M (2008) The evolution of choice policies in UK housing, education and health policy, *Journal of Social Policy*, 38(1): 63-81.

Greve, B (2009a) Can choice in welfare states be equitable? *Social Policy and Administration*, 43(6): 543-556.

Greve, B (2009b) Editorial Introduction, *Social Policy and Administration*, 43(6): 539-542.

Hoxby, C (2003) School choice and school productivity: could school choice be a tide that lifts all boats?, in: Hoxby, C (ed), *The Economics of School Choice*, Chicago IL: University of Chicago Press.

Le Grand, J (2007) *The Other Invisible Hand: Delivery Public Services Through Choice and Competition*, Princeton: Princeton University Press.

Marshall, M (2002) The Publication of Performance Data in the National Health Service, available at: http://www.chi.nhs.uk/eng/ratings/academic_papers/index.shtml .

Paul, S (1992) Accountability in public services: exit, voice and control, *World Development,* 20(7): 1047-1060.

PMSU (2006) The UK government's approach to public service reform – a discussion paper, http://www.strategy.gov.uk/downloads/work_areas/public_service_reform/sj_report.pdf, accessed 15.6.06.

Propper, C, Wilson, D and Burgess, S (2006) Extending choice in health care: the implications of the economic evidence, *Journal of Social Policy*, 35(4): 537-557.

Propper, C and Wilson, D (forthcoming) Competition and choice: the place of markets in connecting information and performance

improvement, in: Walshe, K, Harvey, G, Skelcher, C and Jas, P (eds): *From Knowing to Doing: Connecting Knowledge and Performance in Public Services*, Cambridge: Cambridge University Press.

Schwartz, B (2004) *The Paradox of Choice: Why More is Less*, New York: HarperCollins.

Simmons, R, Powell, M and Greener, I (2009) *The Consumer in Public Services: Choice, Values and Difference*, Bristol: The Policy Press.

Social Market Foundation (2009) *Assertive Citizens: New Relationships in the Public Services*, London: Social Market Foundation, available from http://www.smf.co.uk/assertive-citizens.html, accessed 8.1.10.

Wilson, D, Croxson, B and Atkinson, A (2006) 'What gets measured gets done': headteachers' responses to the English secondary school performance management system, *Policy Studies,* 27(2): 153-171.

Wilson, D (2009) Exit, voice and quality in the English education sector, *Social Policy and Administration*, 43(6): 571-584.