Measuring public value 2: Practical approaches

Dione Hills
Fay Sullivan

Contents

List of Boxes and Tables

Aims of The Work Foundation project

Building on existing academic and policy work around public value, The Work Foundation's project aims to help policymakers, public managers and institutions understand the concept of public value and see how it can be applied in practice.

Public value addresses many of the contemporary concerns facing public managers. These include problems of securing legitimacy for decision making, resource allocation and measuring service outcomes. This research project draws together different strands of the current debate around public value, clarifies its elements and seeks to further understanding of this topical and important conceptual innovation in public service delivery.

The project's objectives are to:
- provide a clear definition of public value
- provide public managers with a set of guiding principles that orient institutions to the creation of public value
- use sector and case studies to illustrate how organisations might understand where gaps occur in achieving public value
- clarify the components and processes of public value in order to facilitate its future capture and measurement.

Sponsors
The project is sponsored by the following organisations:
- BBC
- The Capita Group plc
- Department for Culture, Media and Sport
- Home Office
- London Borough of Lewisham
- Metropolitan Police
- The NHS Institute for Innovation and Improvement (formerly the NHS Modernisation Agency)
- OfCOM
- Quality and Improvement Agency (formerly the Learning and Skills Development Agency)
- Royal Opera House.

About this report
This paper is one of two on measurement. It considers the practical aspects involved in measuring public value and proposes a possible framework for assessing the existing methods by which public value can be measured. Together the two papers examine how public managers are currently measuring public value, the gaps in information, difficulties around decision making, and how a public value framework can resolve these issues.

This paper is one of several background reports being prepared for the public value sponsor group. The research outputs include:

- *Public Value, Politics and Public Management: A literature review*
- *Public Value, Citizen Expectations and User Commitment: A literature review*
- *Public Value and Local Communities: A literature review*
- Sector papers, seminars and presentations on how public value applies to different sectors like local government, policing, skills, broadcasting, arts and culture, and health
- Case studies examining how public value applies to different institutions, ranging from Lancashire Constabulary to the V&A Museum, and as a way of understanding particular local policy issues, such as recycling in Lewisham.

Please note that the views expressed in this report represent those of the authors, and may not necessarily represent those of the project's sponsors.

Executive summary

- This piece of work was funded by The Work Foundation as part of a larger project looking at the notion of public value and its practical utility. It addresses the question of how public value might be measured.

Section 1: Introduction

- The first section of this paper sets out some of the practical challenges that the task of measuring public value brings, such as the mismatch between the notion of measurement, which requires precision and clarity, and the concept of public value, which at this stage remains quite broad and abstract.
- Public value has been heralded as a 'third way' beyond traditional public administration and new public management, but as such it also poses some serious epistemological and ontological challenges. Yet there are also a number of useful areas of work that can contribute to an understanding of how public value might be measured, such as new attempts to measure other broad concepts such as wellbeing, happiness and quality of life, and learning from the field of evaluation about the impact of measurement on that which is being measured.

Section 2: Public value and measurement: Towards a framework of understanding

- The paper then seeks to set out some of the principles that would need to be built into any measure of public value and suggests that this needs to include both what is being measured, and the way in which measures are developed and used, with a key principle being that the measures themselves should contribute to public value.

Section 3: A proposed framework for reviewing measures of public value

- This section introduces a framework that might be used for assessing how appropriate any measure is for measuring public value; for example, is it appropriate, holistic, democratic and trustworthy? The framework is then applied to a number of different systems and measures that have recently been used in the assessment of public service performance to compare how useful and appropriate these are.

Conclusion

- The paper concludes that, while such a framework provides a useful starting point for thinking about how best to measure public value, because public value is still an emerging concept, there is still a great deal of work to be done.

1. Introduction

1.1 The task

The task addressed in this paper is to consider how public value is or can be measured. The paper is part of a larger project that seeks to examine a number of different aspects of the concept of public value and how public value can be applied in practice. The task of considering the measurement of public value was divided into two elements: the economic and the practical issues of measurement. This paper concentrates on the practical aspects.[1] It focuses on how organisations currently measure the value of their services and reviews the literature on measuring value with the aim of ranking the usefulness of some of the practical models for measuring public values.

1.2 The challenge

While this might appear to be a simple task, in practice it proved to be very challenging. The first challenge was a practical one as the term itself is used in relatively small circles. Second, two key search terms, measurement and public value, generated only a small number of references. In these documents there was a number of calls for more work to be undertaken on the measurement of public value and discussion of how the concept of public value could be a useful one in terms of measuring the outcome of public services. Yet there was very little about how this can be done in practice.

By contrast, going beyond the term public value to the literature related to specific measures, such as Best Value Performance Indicators, balanced scorecard, quality of life and wellbeing, opened up an immense literature. This included academic debates about the development and use of such measures, as well as descriptions of specific measures, guidelines on how to use them and IT programmes to facilitate the collection and analysis of information. It was clearly beyond the scope of this relatively limited exercise to review comprehensively such a large literature.

However, a brief review of these publications did indicate that many of them emanate from a somewhat different world from the one in which public value is currently being discussed. One of the features of recent public administrative practices, so-called new public management, has been a strong emphasis on the development of targets and performance measures through which public sector managers aim to manage devolved service provision effectively. This has encouraged the development of many new systems of measurement, sometimes referred to as 'metrics'. Strongly embedded within the world of new public management, these have incorporated similar models and assumptions about the world.

[1] For a discussion of the measurement of public value from an economic perspective, please see Cowling M, *Measuring Public Value: The economic theory*, London, The Work Foundation, 2006

In their analysis of the differences between 'old' public administration, new public management and new public service (akin to the public value approach), Denhardt and Denhardt suggest that these approaches operate according to very different theoretical and epistemological foundations and incorporate different models of human behaviour.[2] (See for example Table 1.)

Table 1: Comparing perspectives: 'Old' public administration, new public management and new public service

	'Old' public administration	New public management	New public service
Primary theoretical and epistemological foundations	Political theory, social and political commentary augmented by naïve social science	Economic theory, more sophisticated dialogue based on positivist social science	Democratic theory, varied approaches to knowledge including positive, interpretive, critical and postmodern
Prevailing rationality and models of human behaviour	Synoptic rationality, 'administrative man'	Technical and economic rationality, 'economic man' or the self-interested decision maker	Strategic rationality, multiple tests of rationality (political, economic, organisational)

While some of these terms need to be treated with caution (for example terms such as positivism and postmodernism have many different interpretations), this analysis does point to some of the tensions that can arise between the worlds of new public management and the one in which measures of public value are sought. Many of the measurement systems currently used can be seen as incorporating a broadly linear model of causality (at least in the way in which they are used) and an empiricist view of evidence (the possibility of objective study of a world 'out there'), for example.

The concept of public value represents a move away from some of these assumptions and models. In public value, causality is complex and multifaceted, evidence is negotiated and contested, the boundary between values and facts is blurred, and the purpose of public service provision is subject to ongoing public debate. There is no inherent conflict between these characteristics and measurement per se, but they do make the whole question of measurement more complex.

Similar challenges have been discussed for some years in the field of evaluation. New approaches such as stakeholder, theory-based and realistic evaluation have been developed to incorporate this more complex view of reality than the positivistic ontology of traditional empiricist methods. However, while these debates are not new to the evaluation world, they are only just beginning to permeate discourses surrounding the management and assessment of public services.

[2] Denhardt R B and Denhardt J V, 'The New Public Service: Serving rather than steering', *Public Administration Review*, Vol 60, No 6, pp549-559, 2000

One reason for the slow transfer of these newer ideas from evaluation to other areas of measurement is that, in accordance with their complex understanding of reality, they have a complex understanding of measurement. This often results in findings that are somewhat less than clear-cut; a trait that may be unsatisfactory to those wanting hard evidence and unambiguous outcome measures on which to make policy decisions. As Blaug, Horner and Lekhi indicate in their literature review of public value, politics and public management, public value has to span the gap between bureaucracy and democracy.[3] This has implications for any measurement system used, which they argue has to address the need for bureaucratic requirements of efficiency and effectiveness at the same time as addressing a democratic need for public accountability. Thus the public value approach seems to be in the awkward and potentially untenable position of requiring both the simplicity and certainty afforded by a positivist ontology and epistemology, and recognising that this conceptualisation of the world is hopelessly inadequate.

This represents one of the central tensions when considering the measurement of public value. It is a tension that kept resurfacing throughout the research and collegial debates that led to this discussion paper and it is one that impedes the offer of a definitive answer to the question 'how do you measure public value?'

1.3 Our approach

We therefore set out to provide a map or framework that would begin the task of navigating this tension and point to issues that needed to be addressed, rather than offer a comprehensive account of tools for measurement.

Our rationale for this was two-fold. First, the lack of any existing measures designed specifically with the measurement of public value in mind meant that our attention was turned to the appropriateness of measures borne out of other paradigms. Second, the nature of measurement requires clear, concrete phenomena that can be measured. However, the world of public value is currently populated with abstract concepts that have not yet been fully operationalised. One of the central features of public value is that this operationalisation of core values is a political task, requiring negotiation between key stakeholders and the public's involvement. It is not a technical task to be left to experts.

We therefore set out to explore the implications of the concept of public value in relation to three questions:

- What needs to be measured in relation to public value?
- How should they be measured, in terms of some suggested guiding principles?
- When might such measures be used?

[3] Blaug R, Horner L and Lekhi R, *Public Value, Politics and Public Management: A literature review*, London, The Work Foundation, 2006

The next section sets out to examine these three questions in more general terms, while Section 3 gives some examples of how these questions might help to identify strengths and weaknesses in existing measurement systems. We also provide an example of how the questions might be used to assess the appropriateness of measurement systems used in a specific sample programme area.

2. Public value and measurement: Towards a framework of understanding

2.1 Introduction

This chapter sets out to provide a map with which to examine the appropriateness of different measurement tools in relation to the concept of public value and in relation to the three questions outlined at the end of the previous chapter:

- What needs to be measured in relation to public value?
- How should they be measured, in terms of some suggested guiding principles?
- When might such measures be used?

2.2 What should be measured in relation to public value?

To the question of what should be measured, one needs to add the question 'what does measurement in the public value approach need to achieve?' Within the public value paradigm, not only do tools need to measure, but they also need to be 'authorised' and generate public value in and of themselves. Kelly et al describe these multiple requirements of measurement in a recent Cabinet Office paper:

> 'Public value provides a broader measure than is conventionally used within the new public management literature, covering outcomes, the means used to deliver them as well as trust and legitimacy. It addresses issues such as equity, ethos and accountability.'[4]

Here, we learn that public value represents a more holistic notion of what public services are for so it will seek to measure a wider variety of issues than other approaches, and that processes are as important as outcomes. Therefore measures are needed both in order to assess the broad outcomes of public services that are implied by the term public value and to assess how well the processes to achieve these are functioning. It also implies, as will be discussed in the next section, that the processes by which measurement is undertaken also need to be examined to ensure that these are contributing to rather than detracting from public value outcomes.

The very term 'public value' points to the fact that a set of core values is at the heart of the assessment of outcomes and the processes by which the outcomes are to be achieved. However, what these values are has not yet been fully defined, in part because they are assumed to be part of the ongoing democratic process through consultation with the public.

2.2.1 Values and valuing

The distinction between 'a value' as a noun and 'what people value' as a verb highlights a central confusion that appears to run through discussions about public value. The notion of a value (as a noun) usually refers to broad, abstract qualities or principles that might be derived from various sources, eg moral or ethical debate, political or personal commitments. As such, they represent a set

[4] Kelly G, Muers S and Mulgan G, *Creating Public Value: An analytical framework for public service reform*, London, Cabinet Office, 2002

of standards against which something is compared, although they also usually represent a direction rather than a destination. For example, it is unlikely that perfect equity, safety, quality of life and so on will be achieved.

The notion of what people value is rather different and operates more in the realms of opinion and satisfaction scales – an expression of preference rather than comparison with a more objective standard. For example, each person interviewed in a satisfaction survey could be assessing their satisfaction with a service according to a different set of values. However, the overall result is concerned with the overall level of satisfaction measured, rather than the values individuals are using to make their assessment.

The distinction was highlighted in an exercise undertaken by the King's Fund, which sought to identify a set of key values that might underlie the NHS.[5] The team involved made a distinction between what they describe as 'moral' and 'non-moral' values. Non-moral values represent the more subjective element of personal tastes and preferences. Moral values are not purely a matter of personal preference or individual assessments of goodness, desirability and rightness, but judgements about how people ought to behave generally, irrespective of personal preference. Staley underscores this distinction: 'In the context of public debates, it is important to keep a clear distinction between "public values" and the "public's values". The former are moral concepts, the latter are empirical findings about what position people take on those moral concepts.'[6]

In terms of these distinctions, Staley goes on to give a very specific example of this:

> 'For example, in relation to public health policy that aims to reduce smoking, a debate on public values would revolve around a discussion of the limits to government action to improve health through smoking cessation. It would consider normative issues such as "would a ban on smoking in public places be acceptable?" And "is it acceptable to place a heavy tax on cigarettes?" In contrast, the public's values are simply what the public thinks about these issues at any given point in time.'[7]

An investigation of the discourse about NHS values in official documents produced a catalogue of eight core 'meta' or higher values, which included health, universalism, equity, democracy, choice, respect for human dignity, public service and efficiency.[8] The project identified two key difficulties with the concept of core values of this kind that are particularly relevant to the discussion of measurement. The first was that while such higher-level values were useful in winning support

[5] Appleby A, Harrison A and Devlin N, *What is the Real Cost of More Patient Choice?*, London, King's Fund, 2003
[6] Staley cited in New B and Neuberger J (eds), *Hidden Assets: Values and decision making in the NHS,* London, King's Fund, 2002
[7] Ibid
[8] New B and Neuberger J (eds), Ibid

across the NHS, for all practical purposes each profession behaves according to its own particular core values and beliefs. In this respect, Hunt argues that such values have a useful symbolic function for public consumption, but did not serve particularly well in terms of providing a corporate framework or bond for NHS staff.[9] Translating them into specific services highlighted the second problem – the fact that some of these values could, in practice, work against one another. Seeking to be democratic could undercut efficiency and allowing choice and being fully equitable could undercut universalism. Even if it is possible to reach agreement in a diverse population on higher-level notions of 'equity' or 'efficiency', the same concepts can be highly contested when operationalised into specific activities or in specific contexts, particularly in terms of which values should take priority.

This provides a very concrete example of how the translation of broad values into concrete, measurable practice brings to the surface difficulties that might previously have been overlooked. In terms of the measurability of public value, it also draws attention to the fact that the values themselves by which the processes and outcomes of public services should be judged have yet to be decided.

2.2.2 The core values in public value

As noted before, the process of deciding key values is part of the public value process, and which values are viewed as particularly central are likely to vary over time and according to which sector or policy area is being considered.

Having said that, it is possible to detect from the literature to date 'clusters' of values that are being discussed. Some of these overlap or blend into one another. For example, there are clusters relating to the process of public service delivery. These include the new public management values of efficiency, effectiveness and cost effectiveness as well as broader values such as:

- democracy (involving the public)
- transparency
- equity
- authorisation: negotiation between different stakeholders
- trust.

There are also clusters of values that relate to the outcome of public services that go beyond the delivery of specific service outcomes to include notions of enhancing:

- quality of life, wellbeing and happiness
- social capital, social cohesion and social inclusion
- safety and security – involves the subjective experience of safety on the part of the public, as well as actual freedom from crime and attack

[9] Hunt cited in Ibid

- equality, tackling deprivation and social exclusion
- and promoting democracy and civic engagement.

There are probably other values of relevance here for particular sectors. For example, enhancing cultural cohesion has been discussed in relation to the public's funding of broadcast media. It is likely that other core values will emerge as the notion of public value is applied to different areas of policy. The challenge of developing and using measurement scales in relation to these 'clusters' is discussed in Section 3.

2.3 How?: the process of measurement

The process of measurement, as already indicated, is a precise and often complex task that requires the operationalisation of broad concepts into specific measurable objectives or behaviours. Answering the question of how public value can be measured draws attention not only to the kind of measures that are needed, but also to the processes by which the measures are developed and used.

This is because, as already noted, the notion of public value itself draws attention to the fact that it is the processes as well as the outcomes that are important in considering how public value is created because it is these processes that contribute to, or undermine, this goal. In Mark Moore's terms, the authorisation, delivery and task systems are all equally important.[10] In the terms of Kelly et al's paper, the quality of the service as well as the broader outcomes achieved and the trust generated are all central to public value.[11]

The importance of the processes involved in the development and use of measures is indicated by a growing literature about the way in which the use of performance measures and how these are reported to the public can contribute to or undermine public confidence in the quality of public services.[12] This suggests the need for a set of standards or criteria by which measures, their development and use can be assessed in order to ensure that they contribute to, rather than undermine, public value.

After a discussion of the challenge of measurement in relation to public value, the following sections suggest some of the criteria or standards that should be applied.

2.3.1 The challenge of measurement

Measurement is one of the principles of science as laid down by Isaac Newton, which required the accurate observation of nature and systematic development

[10] Moore M, *Creating Public Value: Strategic management in government*, Cambridge, MA, Harvard University Press, 1995

[11] Kelly G, Muers S and Mulgan G, *Creating Public Value: An analytical framework for public service reform*, London, Cabinet Office, 2002

[12] Wilmot A, Jones J, Dewar A, Betts P, Harper R and Simmons E, *Public Confidence in Official Statistics: A qualitative study on behalf of the Office for National Statistics and the Statistics Commission*, London, Office for National Statistics, 2005

and testing of hypotheses. This required a level of precision in terms of how measures were taken, as well as a level of rigour in relation to the development of theory and generation of hypotheses.

At the level of measurement, scientific progress was achieved in part through the development of sophisticated measurement technology that enabled primary qualities of objects such as weight, size, temperature and density to be measured accurately and reliably. The ontology underlying mainstream science is (usually) one in which reality exists and is driven by laws of cause and effect, and its epistemology is one in which this reality can be determined through empirical study free from subjective judgement.

However, the application of these principles to the social world has always been rather more complex and contested, partly because this involves the study of subjective qualities (such as pain or satisfaction) and broad abstract notions such as intelligence or democracy.

In terms of subjective values, early scientists followed seventeenth-century philosopher John Locke's solution of distinguishing between primary (objectively measurable) qualities such as height or weight, from secondary qualities such as taste or colour, which existed only as part of human experience. Unfortunately, such a neat distinction was soon undermined by George Berkeley, who argued that all experience is in fact phenomenal, limited to appearances in the mind and thus, in some respects, subjective. Later philosophers added to the difficulty by pointing out the centrality of language and conceptual frameworks to the way in which phenomena are perceived. This posed a particular challenge in the study of social phenomena, since it was clear that those studying the social world were themselves closely embedded in it and would have great difficulty in bringing 'objectivity' to bare.

The very possibility of further objectivity was challenged by the emergence of interpretist, hermeneutic and, more recently, varieties of postmodern schools of thought. Most of these schools challenge the notion of a 'reality' out there beyond the constructs and meanings that are ascribed to it, and place considerable attention on the question of who is undertaking the study, and what assumptions and models they bring to the task. Some of these schools of thought have been influential in the field of evaluation, where attempts have been made to develop evaluation strategies that incorporate the idea that social phenomena are negotiated and political processes can determine those aspects that are emphasised or ignored.

Broad abstract or, in the case of values, potentially metaphysical notions drawn from morality, ethics or religious belief have presented a different challenge to scientists, particularly since much of early science was devoted to testing and dismissed metaphysical explanations of the natural world. In the early twentieth century, the logical positivists of the Vienna circle asserted the meaninglessness of all metaphysical statements, which they argued were based on propositions that were neither verifiable by empirical observation nor demonstrable as analytic. This hard line has now generally given way to a more muted version in logical empiricism, which focuses more on constructing logically explicit and rigorous concepts and theories that can be tested through empirical (and particularly quantitative) study. This view has been widely influential in the development of the social sciences.

Much effort has since been expended in the development of methodologies and measurement scales through which broad and essentially abstract constructs related to social phenomena can be studied as rigorously and objectively as possible. However, this involves considerable technical expertise as well as complex statistical practices, and this means that any implicit models or theories embedded in these methods or measurement scales can be hard for the layperson to detect.

This can be a particular difficulty because due to the cultural authority ascribed to science and technology, such scales can come to take on a cultural life of their own, awarded a similar 'reality' value to the phenomenon they seek to measure. For example, 'intelligence' can often be seen as synonymous with a measure developed to test it – the IQ test. At a more profound level, the very notion that there is 'normal' behaviour is derived from the statistical study of populations:

> 'Nineteenth-century medical men were developing a language for situating all people in relation to each other, for measuring their deviation from the normal, and increasingly for managing their deviations from that norm. The normal began to be seen as the expression of natural law.'[13]

This places an added emphasis on the task of ensuring that measures of social phenomena are accurate, valid and reliable, which relies on ensuring that the processes of developing these are as transparent as possible. Although considerable efforts are made to ensure that this is the case, it is very easy for major assumptions to be built into the heart of such measures on transparency, as will be discussed in the section below.

2.3.2 A set of standards for the measurement of public value

The discussion above highlights the importance of having standards or criteria by which measures and the way in which they are developed are judged.

[13] Lawrence C, *Medicine and the Making of Modern Britain 1700-1920*, New York, Routledge, 1994

Traditionally, the standards applied to the development of measures have been those emphasised by scientific practice – standards such as accuracy, validity and reliability.

However, the notion of public value suggests that there are some wider principles or standards that need to be taken into account in order to ensure that the measures developed are conducive with a wider set of values. For example, it indicates principles such as:

- appropriate measurement
- democracy (involving the public)
- transparency
- complexity (ensuring that the complex nature of modern public services are fully taken into account)
- negotiation between different stakeholders
- trust.

2.3.2.1 Appropriate measurement

Assessing whether a measure is appropriate involves many traditional standards: ensuring that any measures used are valid, objective and reliable etc. In particular, it involves ensuring that the measures used do reflect the phenomenon being measured. This can be particularly difficult when the phenomenon being assessed is a broad, abstract one that might be captured by the measure of a limited set of attributes. For example, measures of equality might involve keeping a count of numbers using a service who have particular attributes, eg ethnic background, age, gender, level of disability. However, it might exclude more subtle causes of inequality, such as discrimination against people from a particular housing estate or cultural differences as expressed in ways of speaking or dressing.

Ensuring that measures are appropriate requires clarity about how what is being measured relates to the purpose to which the measures are being used. This requires considerable work in ensuring that there is clarity in the underlying model about how a particular policy or programme is intended to achieve its objectives. For example, clarity about the underlying 'programme theory' or model in a programme designed to reduce crime through the appointment of neighbourhood wardens and increased CCTV can help identify the measures used in monitoring and whether the programme has been effectively implemented – as well as what outcome and impact measures would be most appropriate.

In the evaluation field, there has been concern for some years that earlier simplistic models of evaluation took for granted both the appropriateness of the original assumptions underlying a policy or programme and whether it had been fully implemented. Another frustration with early approaches to evaluation, particularly those incorporating experimental methods, was their failure to take into account the context in which an intervention took place and its potential influence over the outcome.

This has led to the development of more complex models of evaluation in which outcome measurement is linked to a more sophisticated understanding of the model or theory underlying the programme[14], and a greater interest in gathering data about the context in which it is implemented. For example, theory-based evaluation starts from the assumption that it is important to understand what it is about a programme that makes it work (studying processes), rather than just knowing whether it works or not (ie measuring outcomes and impacts).

One particular model of theory-based evaluation – realistic evaluation – proposes from a scientific realism position that causal explanations are based on generative rather than successionist principles (cause describes the transformative potential of phenomena: one happening may well trigger another, but only if it is in the right conditions in the right circumstances).[15] The classic example is that a spark is intended to bring about an explosion when applied to gunpowder, but it cannot do this if the gunpowder is damp. Pawson and Tilley suggest that the key question in any evaluation is therefore not just 'what works?', but 'what works for whom under what circumstances?', with the central focus being studying the mechanisms through which the interventions are intended to bring about a change in the underlying circumstances.[16] They argue that without understanding these, it is unclear exactly what the key factors are in terms of the change achieved, and what it is that should be replicated if the intervention is to be repeated.

These developments in the evaluation field help to highlight the fact that the 'appropriateness' of measures used in the assessment of policies and programmes rely not only on their intrinsic validity, but also on the models underpinning the way in which they are being used. So, the use of a measurement system using a linear or successionist model of reality may be inappropriate in circumstances in which the link between cause and effect is complex and contested.

2.3.2.2 Democracy

The notion of public value places a strong emphasis on the importance of involving the public in setting the values that will determine the choice of services, as well as in providing feedback on the effectiveness of services in achieving their aims. However, there is a real tension between the technical skills required in the development of measurement systems that fully reflect the requirements of accuracy and validity, and the desire for public involvement.

---

[14] Chen T H and Rossi P H, 'The Multi-Goal, Theory-Driven Approach to Evaluation: A model linking basic and applied social science', *Evaluation Studies Review*, Annual 6, pp38-54, 1981
[15] Pawson R and Tilley N, *Realistic Evaluation*, London, Sage Publications Ltd, 1997
[16] Ibid

This can be a real challenge. For example, part of the task undertaken by the King's Fund was to gather empirical data about the public's views on the sets of values they had identified as being central to the NHS. This proved to be far from easy, because it required people to engage with one another about values that represented something more than their personal interests. More widely in the health field, attempts to involve patients or the public in the planning and delivery of research programmes has been resisted on the grounds that these issues are too complex for most patients to engage with effectively.

In spite of this, real attempts have been made to involve the public in the process of developing and using measures. For example, there is now a national advisory group, 'Involve', funded by the Department of Health, that has been set up to promote and support active public involvement in NHS, public health and social care research. This group has highlighted the difficulty of public involvement being little more than a sham exercise unless considerable effort is made to ensure that those involved receive sufficient training and support. To this end, the National Institute for Health and Clinical Excellence (NICE) has allocated considerable resources to recruiting and supporting patient representatives in assessment panels for new clinical procedures.

Underlying the discussion of how best to involve the public in setting standards and developing measures are some very different notions of democracy. In traditional public administration, the measurement of public views was undertaken through representative democracy, in which publicly elected representatives made decisions. However, more recently there has been more emphasis on direct democracy, where attempts are made to assess how the public 'feels' about a particular issue via some kind of consultative process or public opinion polls and satisfaction surveys. This use of measurement will often assume that the view of the majority (as in representative democracy) provides the answer to the question of what the public wants or values. However, this overlooks some major divisions that might exist. For example, different sectors of the public might give different views arising from different levels of knowledge.

Some research methods such as focus groups, Delphi techniques and qualitative investigations have been developed to try to go beyond any simple 'majority' conclusions and create opportunities for debate and discussion. It could be argued that these reflect the concept of deliberative rather than representative democracy. Other attempts to address the issue of mediation of different viewpoints are discussed in Section 2.3.2.5 on negotiation.

2.3.2.3 Transparency

In order to achieve effective involvement from either the public or other stakeholders in the development and use of measurements, there has to be transparency in terms of what, or whose, assumptions have been built into any measures used.

This can be hard to achieve given the technical nature of many measurement systems. For example, even the simplest numerical device – a nominal scale in which people are classified into categories such as male/female, married, divorced, single etc and counted – has ambiguity and complexity that can be ignored or glossed over. For example, such scales ignore the fact that there may be individuals who cannot be classified neatly as either male or female, or those whose marital status is far from clear.

The more sophisticated the scale, the more opportunities there are for assumptions to be built in. Ordinal scales are those in which phenomena are ranked according to some quality, such as whether they are seen as bigger, better or more effective than others. A star-rating scale is an example of an ordinal scale that may place considerable emphasis on one aspect or set of aspects being rated. In interval or ration scales, real numeric values are ascribed to phenomena that allow for more sophisticated statistical analysis. This might be relatively straightforward, for example in ascribing a numerical value to an individual's age. However, in most cases the establishment of a single numeric scale requires the combination of a number of different attributes, such as various aspects of health (sense of wellbeing, functional ability, level of discomfort or pain) into a single measure of 'health'.

The term 'metrics' is increasingly used to refer to the process of assigning numerical values to phenomena of this kind, particularly where these involve the combination of a number of different elements together into an overall numeric scale. A number of difficulties can creep into attempts to aggregate scores in this way. For example, when scores are aggregated across multiple sites, as in combining test results from different schools or crime figures from different police stations, these may fail to reflect local conditions that have affected the way they are collected.

Another difficulty is that in most such scales, some of the variables have been weighted in ways that are not apparent to the outsider. Who determines the weighting and on what criteria, and what values this weighting embodies, are crucial to the overall result. In some cases, the sample on which the weighting was based might be very small. For example, the Rosser index, a health scale widely used in clinical trials that combines measures of disability and distress, was drawn up using weightings derived from a sample of only 70 subjects, including doctors, nurses, psychiatric patients and healthy volunteers.[17]

Transparency would require much greater clarity, for example about who has developed measurement scales, how many (and who) were involved in the testing, and particularly with aggregate scales, who has been involved in setting any weightings and on what basis.

2.3.2.4 Complexity

One of the features mentioned throughout the literature on public value is that it represents a world in which services are delivered through a range of different agencies, some of which might be relatively autonomous. Additional complexity is built in through the involvement of multiple partners in service delivery and in the growing realisation that many of the desired 'outcomes' of public service provision are affected by multiple causes. For example, it is now widely recognised that a wide range of factors – including poverty, unemployment, housing and education, as well as any factors over which a health service has direct control – influences health and wellbeing.[18]

This means that the use of measurement, in particular in the assessment of outcomes and service impacts, has to go beyond any simple, linear notions of cause and effect and begin to embrace complexity. This is beginning to be taken into account in the analysis of social problems, as in the application of socio-ecological models to the understanding of health inequalities. There is also a growing awareness that complexity and chaos theory may have a role in evaluating some publicly funded interventions.[19] However, at present the actual application of such models and approaches is very limited, in part because the systems of measurement and mathematics involved are highly complex and outside the scope of many of those involved in the provision or assessment of public services.

2.3.2.5 Negotiation between stakeholders

The concept of public value highlights the fact that the authorisation, delivery and performance of public services involve multiple actors and, with the ending of simplistic notions of hierarchical provision, requires negotiation between different interests. This requires that there are spaces and opportunities in which such negotiations can take place as well as processes, hopefully, in which different levels of power and influence are mediated.

The importance of negotiation is highlighted by the King's Fund exercise mentioned earlier in which eight core NHS 'values' were identified, but which also found that in any given situation these values were often in conflict. Typical

[17] Rosser R, 'A Health Index and Output Measure' in Walker S R and Rosser R M (eds), *Quality of Life: Assessment and application*, Lancaster, MTP Press, 1988
[18] Acheson Sir D, *Independent Inquiry into Inequalities in Health Report*, London, HMSO, 1998
[19] Sanderson I, 'Evaluation in Complex Systems', *Evaluation*, Vol 6 No 4, pp433-54, 2001

conflicts were identified between choice and equity, equity and efficiency, democracy and equity, and efficiency and universalism. It was often unclear how such conflicts of value should be resolved.

There may also be tensions between the different bodies that are consulted as representatives of the public. For example, in terms of public feedback on services, it is now widely recognised that greater satisfaction is recorded for those who are direct users of the service than for those who are not direct users. Different views might also be put forward by groups representing different sets of interest. For example, in an Oregon project that involved consulting the public on a key set of medical procedures and conditions that should be funded out of Medicaid, it was disability groups who contested the use of quality-of-life scales in the overall assessment process.

In traditional approaches to measurement there has rarely been any opportunity for different stakeholder views to be acknowledged, let alone any space provided for negotiation between these views. In part, this is because in most uses of measurement (as in performance measures) there is an assumption (what might sometimes be referred to as the positivist assumption) that there is one 'reality' that the measure is designed to uncover rather than multiple 'realities' to be mediated. However, in the wider field of evaluation this issue has been addressed regularly. Indeed, it is now a standard aspect of good evaluation practice that the views of all stakeholders should be sought and taken into account (eg American Evaluation Society good practice standards).

Guba and Lincoln led the way in relation to stakeholder evaluation and sought to apply the principles of constructionist, hermeneutic and interpretist philosophies of science to the practice of evaluation.[20] These argue that rather than there being one objective reality that can be observed, realities exist as mental constructs and are relative to those who hold them, with knowledge and research findings representing the result of an interaction between researcher and researched. This has led to questioning of earlier calls for objectivity on the part of the researcher, arguing that initiatives and 'findings' are seen to be the result of negotiation between different stakeholders. In some cases, the evaluation process may be a major forum in which such negotiations take place, with the evaluator as the 'orchestrator' of a negotiation process.

Although quite widely debated in the evaluation community, such views of the world are still found to be somewhat uncomfortable by those working in the field of developing and using measures and metrics.

---

[20] Guba E and Lincoln Y, *Fourth Generation Evaluation*, Newbury Park, California, Sage Publications, 1989

2.3.2.6 Trust

The importance of public services being trustworthy is one of the central features of the model of public value put forward by The Work Foundation.[21]

In some respects, the incorporation of some of the principles outlined above – of transparency, accuracy, democracy and negotiation – in the development of measures can help to ensure that these are seen to be 'trustworthy'. However, how systems of measurement are used can also be central to their trustworthiness. In other words, a good measure might be used in ways that undermine trust.

This brings in the whole question of how performance measures might undermine the quality and performance of the very service they seek to measure; something that Ridgeway noted in 1956 when he wrote about dysfunctional aspects of performance measurement in the first volume of *Administrative Science Quarterly*.[22] This situation can arise from various sources, including cost (in terms of staff time, computer resources) as well as more hidden difficulties in terms of distorting the way in which services themselves are provided. Smith made a systematic study of some of these less apparent unintended consequences:

- **Tunnel vision:** service managers choose targets that are easiest to measure and ignore the rest
- **Sub-optimisation:** service managers operate in ways that serve their own operation, but damage the performance of the overall system
- **Myopia:** managers focus on short-term targets at the expense of longer term objectives
- **Measure fixation:** a tendency to focus on the performance indicator itself, rather than the desired outcome
- **Misrepresentation:** where performance data is either misreported or distorted to create a good impression
- **Misinterpretation:** where there are no real differences between units that appear at different points in a league table
- **Gaming:** when a manager deliberately underachieves in order to secure a lower target next time around
- **Ossification:** when performance indicators are no longer measuring anything meaningful, but have not been revised or removed.[23]

Although some perverse effects are deliberate due to deliberate misrepresentation or gaming, others occur despite the best efforts and good intentions of those developing the measuring systems. A study of UK local authorities being inspected by the Audit Commission as part of the government Best Value system identified a number of ways in which distortions occurred,

---

[21] See for example Horner L, Blaug R and Lekhi R, *Public Value Final Report: Deliberative democracy and the role of public managers*, London, The Work Foundation, 2006

[22] Ridgeway V F, 'Dysfunctional Consequences of Performance Measurements', *Administrative Science Quarterly,* Vol 1, pp24-247, 1956

[23] Smith P, 'On the Unintended Consequences of Publishing Performance Data in the Public Sector', *International Journal of Public Administration*, No 18 Vols 2 and 3, pp277-310, 1995

which are referred to as 'virtualism'.[24] For example, pursuit of high scores on performance indicators can lead to the exaggeration of hard quantitative over soft qualitative data.

It was partly in response to what appeared to be a declining level of trust in public institutions identified by the Cabinet Office[25] that the Office for National Statistics undertook a recent study of public trust in official statistics.[26] This found that trust in government statistics was closely related to education and familiarity with the use of statistics. More highly educated groups tended to have a more informed view of statistics, while less highly educated groups related the discussion more closely to their own personal experience or that of people they knew. Trust in the independence and accuracy of official statistics was influenced by whether or not people perceived that their production was free from political interference, as some thought that those with a vested interest in the results could influence the production process at any stage. Distrust was also based on confusion over the fact that different statistics appeared to be produced from the same sources, or different sources produced different statistics on the same subject (eg police statistics on reported crime versus levels of crime reported in the annual crime survey).

2.4 Putting together what is measured with how

Putting together the list of outcome measures outlined in Section 2.2 with the discussion above about the criteria for measures that would promote rather than undermine public value begins to provide a map of how different measurements would look in a public value regime. The shaded area in Table 2 overleaf indicates the areas that have received most attention in a new public management framework.

The shaded area would still need attention in a public value framework, but the other boxes would also require attention giving a much wider framework. Thus, measures related to efficiency, effectiveness and cost effectiveness would still be needed, but greater attention would need to be paid to whether there was public involvement in these, whether they were transparent and fully addressed the complexity of the policy or programme. There would also need to be space for negotiation between stakeholders about the appropriateness of the measures chosen to ensure that these promoted rather than undermined trust in public service provision. The same questions would need to be addressed to any measures chosen to assess wider impacts such as the contribution that the policy or programme was making to quality of life, social capital or civic engagement.

---

[24] Miller D and Carrier J G, *Virtualism: A new political economy*, New York, Berg Publishers, 1998

[25] Cabinet Office, *Political Trust and Engagement Seminar Papers: Public assessment and performance*, unpublished, 2004 and quoted in Wilmot et al, 2005

[26] Wilmot A, Jones J, Dewar A, Betts P, Harper R and Simmons E, *Public Confidence in Official Statistics: A qualitative study on behalf of the Office for National Statistics and the Statistics Commission*, London, Office for National Statistics, 2005

Table 2: Questions concerning methods of measurement

Questions concerning methods of measurement		What is measured		
		Process measures: effectiveness and efficiency	**Policy and programme-specific outcome measures**	**Broader outcome and impact measures: quality of life, safety, civic engagement etc**
Is the method:	**Appropriate?** • Are they fit for purpose? • Do they meet relevant methodological standards, including method-specific quality standards?			
	Holistic? Do they take into account: • the complexity of the situation? • new public management values such as effectiveness and efficiency? • relevant public value-type values, such as wellbeing, social capital and quality of life?			
	Democratic? Do they allow for: • public involvement? • negotiation between different stakeholders? Are they transparent and accessible? Have they been 'authorised'?			
	Trustworthy? Are the measures and their findings being used appropriately and with integrity?			
	Generating public value? Does the process of measuring create value in and of itself?			

It is interesting to note that the Atkinson review highlighted a number of similar issues to the ones that we have identified in relation to finding measures for assessing the contribution of spending on public services in terms of productivity.[27] The review put forward a set of nine principles that need to be incorporated into the development of measures, including:

- **Complexity:** Any change in productivity is extremely complicated to measure, with no single number, however carefully constructed, fully capturing the performance of complex public services with multiple objectives. The review highlighted the importance of using a range of information and triangulating information
- **Holistic:** Current indicators have been too limited in their coverage of activities and have been aggregated at too high a level
- **Transparency:** 'We urge [the] ONS and the departments to be transparent…It would be highly regrettable if the objective study of a matter of public importance were to be inhibited by misunderstandings and public criticism of figures that are clearly interim.'
- **Measurement of quality:** Quality has many dimensions and some will prove illusive. Measurements should be representative of the range of dimensions. This is not straightforward and may take some time.[28]

2.5 When measures are used: policy and programme cycles

Different kinds of measures are required at different stages of the planning and delivery of a programme. In order to facilitate discussion about the kinds of measurement required at different points in the cycle, it is helpful to clarify some of the different stages in policy and programme cycles.

For example, the EU's guide to social policy evaluation breaks the policy cycle down into the following stages:

- needs assessment/agenda setting
- planning/policy formulation
- policy implementation
- policy monitoring
- evaluation and feedback.

Similarly, programme cycles are broken down into:

- agenda setting
- planning/programme formulation
- programme implementation
- programme monitoring
- evaluation and feedback.

Of course these are not just linear stages, but points in a cycle, with evaluation and feedback feeding back into needs assessment and planning.

[27] Atkinson T, *Final Report of the Review of the Measurement of Government Output and Productivity for the National Accounts*, London, ONS, 2005
[28] Ibid

For the present task, in order to keep our model as generic as possible, we are using the more general terms of:

- needs assessment/agenda setting
- task specification
- delivery/implementation
- outcomes (specific to the policy and programme)
- broader impacts (particularly in terms of broader public values outlined earlier).

Evaluation processes used at different points in this cycle are generally ex anti evaluation, ongoing or process evaluation (also monitoring and audit), impact or ex post evaluation, as Table 3 shows.

Table 3: Type of evaluation process by cycle point

Stages in policy or programme cycle	**Needs assessment/ agenda setting**	**Task specification**	**Delivery/ implementation**	**Outcomes**	**Impacts**
Type of evaluation	Ex anti	ongoing or process		impact or ex post	

2.5.1 What is being measured when

Clarifying the different stages of a policy or programme helps to identify what different types of measures are required at different stages, and how this might be different from a public value perspective (see Table 4 overleaf).

In terms of the different stages identified, it is important to remember that measurement is only one aspect of these processes of evaluation, assessment and review. These also generally involve various other forms of consultation, debate and discussion, with measurement systems as outlined above one part of or feeding into these processes. However, one of the difficulties that has sometimes arisen around measurement systems is that the factors that can be measured have too much influence over the whole process of evaluation and review. For example, this happens when performance measures that relate to only limited aspects of a programme are the main criteria by which a programme or policy are judged. Another difficulty that has been noted is when policies are decided based on available evidence, even when only a limited range of the various options have been subjected to systematic study. (An example of this might be the establishment of a smoking cessation programme based on a relatively small number of interventions that have been subjected to experimental study.)

Table 4: Type of measurement by stage

Measurement at the planning stage	
New public management	**Public value**
Needs assessment – surveys, use of existing statistics about populations (eg unemployment data)	Assessment of acceptability of policy options – may involve opinion polls or surveys
Assessment of the evidence base for different options	Examining evidence that the policy can contribute to wider public value outcomes and impacts
Assessment of the relative cost of different options – as recommended by HM Treasury's *Green Book*[29]	
Measurement at the delivery stage	
Monitoring use of resources, staff involved, number of services provided to number of recipients	Ensuring monitoring addresses equity issues
Provision auditing – either in terms of cost or in a more generic way in terms of assessing actual provision against a clear set of criteria	Ensuring criteria for assessment incorporate wider values such as equity
Ongoing or developmental evaluation – evaluation information is gathered on various aspects of the delivery process and fed back to enable changes and improvements to be made	Ensuring full range of stakeholders involved in development of evaluation process
Measurement at the outcome and impact stage	
Performance measurement in terms of hoped-for outcomes, often using monitoring data or other statistical data gathered as part of routine administrative processes	Ensuring performance measurements are appropriate
Outcome and impact evaluation – specific tools are developed to capture information about outcomes and impacts	Ensuring wider public value impacts are incorporated into the evaluation
Satisfaction measures – set up to assess service users' satisfaction with the service provided	Ensuring these incorporate the full range of service users and possibly also broader public assessment of outcomes

If performance measures are to be a key criterion by which a programme or policy is judged, then the selection of appropriate performance measures is therefore crucial. In this respect, it is likely to be particularly important to refer to the checklist of values identified earlier in Section 2.2.2 to see if some of the broader values are applicable. However, particularly in the case of these broader values, it will be relevant to ensure that expectations are realistic, ie that the inputs are those that will achieve these outputs.

Two dimensions often lacking both in conventional performance measurement and particularly important if broader impacts are being considered are the scale

[29] HM Treasury, *The Green Book: Appraisal and evaluation in central government*, London, HM Treasury, 2003

of impact anticipated and over what time period. An example of this can be drawn from the New Deal community programme. Evaluation results indicated that partnership activities were often successful, but there was little evidence of the achievement of any measurable change in the unemployment or crime rates in target communities.[30] Some saw this as an indication of the programme's failure, although the initial conceptualisation appeared to involve little theoretical consideration of what scale of change might be achievable by such measures and over what time period.

One further point relates to the role of measurement at different stages of a programme and how this might contribute to both creation and assessment of public value. We suggested in Section 2.2 that measurement required a level of clarity and specificity that may be lacking from early-stage discussions of wider values. As the King's Fund project indicated, translating broad values into specific services can highlight differences in views about how they are to be operationalised, as well as potential conflicts between different sets of values.

In terms of the policy and programme cycle, we would also like to suggest that the measurement process will often serve to highlight ambiguities or problems as a policy moves from the planning stage through to implementation and outcomes. These difficulties will usually reflect a problem that has not been fully resolved, or a conflict of interest that has not been successfully negotiated at an earlier stage. In the example of the New Deal Community programme, the lack of evidence of impact in the anticipated measurements might lead to reviewing initial expectations about the timescale in which such results might have been achieved, or whether the scale of change anticipated was unrealistic given the design of the programme – that it was too early in the process to detect the kind of long-term change the programme hoped to achieve. It might also help to pinpoint an absence of sufficient negotiation between those setting up the programme on the ground and the centre, or between the centre and the evaluation team. In either respect, the failure to achieve anticipated outcomes can be seen to represent a lack of clarity about how the public value that the programme was expected to produce was to be operationalised, and exactly what measures were the most appropriate in terms of capturing this value in the short, medium and longer term. Thus the measurement process can potentially contribute to public value, not just through demonstrating that the programme or policy as a whole is faulty, but in drawing attention to the need to re-examine decisions made further up the policy/planning chain. This can help to ensure that similar mistakes in terms of public value outcomes are not repeated further along the cycle or in subsequent policies or programmes.

[30] Neighbourhood Renewal Unit, *Research Report 7: New Deal Communities, National evaluation – Annual Report 2003-2004*, London, DCLG, 2003

2.6 Conclusions

In this section we have attempted to outline some of the key elements that need to be examined when considering the measurement of public value. We highlighted the importance, in terms of measurement, of gaining greater clarity about what the key values were in relation to public value, particularly in terms of the outcomes that were sought. These would indicate what measurement systems should focus on. We then identified some of the values that might help to ensure that, if applied to the practice of developing and using measurement systems, these contributed to rather than undermined public value in the services in which they were used.

Putting these two aspects of measurement together helped to indicate how measurement in the context of public value might look different from current systems of measurement, for example as used in performance measurement. We then outlined the different stages in policy and programme development in which measurement might be used, how different measures are used at different stages and how these might be different in terms of public value. We also suggested that the concept of public value might lead to different ways of thinking about measurement over the course of a programme. For example, it might help identify the need for a clear conceptualisation of how different inputs or approaches to the delivery of the programme are intended to lead to measurable outcomes. In turn, failure to achieve anticipated outcomes might encourage a review of earlier stages in the planning and delivery cycle and identify where there was a lack of clarity or where there had been insufficient negotiation over what it was hoped that the policy or programme would achieve, and over what time scale.

The next chapter examines some existing measurement systems and programmes using some of the ideas put forward in this chapter.

3. A proposed framework for reviewing measures of public value

3.1 Why develop a framework?
A comprehensive notion of public value is still really in a fledgling state. A literature is beginning to grow around the concept, but it is positioned at the level of the abstract rather than the practical and still concerned with establishing a shared definition and understanding of public value. Very little work has been undertaken around the actual measurement of public value. The task of measurement is a highly practical one and attempting it for the public value approach highlights that there remains many complex questions to be answered about the approach in general. As the literature does not yet contain any methods developed specifically for the measurement of public value, we turned our attention to existing methods of measurement and asked whether or not they would be appropriate for use in a public value paradigm. We developed the framework outlined below in Section 3.2 to help answer this question.

3.2 The framework
The framework is generic in that it is capable of being applied to any method of measurement, regardless of its ontology and epistemology or the scientific paradigm in which it originated. Running down the left-hand side of the tabular representation of the framework (see Table 5 overleaf) is a list of questions to be asked of or criteria that should be met by the method under consideration if it is to function in the public value paradigm. Measures should be appropriate, holistic, democratic, trustworthy and their application should generate public value.

Measures should be appropriate, fit for purpose and meet relevant methodological standards, including method-specific quality standards. This set of criteria should be met by all measures regardless of whether or not they are being applied in a public value approach. What public value does here is remind us of the implications of not meeting these criteria. Failure will not simply result in internal and external invalidity or inaccuracy, but it may also lead to perverse effects and a lack of trust in the methods, its findings and the people who use them. In short, failure may destroy public value. Good-quality standards that can be applied to a wide range of methods include *The Green Book, The Magenta Book, Quality in Qualitative Evaluation* and *Types and Quality of Knowledge in Social Care*.[31] Quality standards specific to the method should also be adhered to.

[31] See HM Treasury, *The Green Book: Appraisal and evaluation in central government*, London, HM Treasury, 2003; Government Chief Social Researcher's Office, *The Magenta Book: Guidance notes on policy evaluation and analysis*, London, Government Chief Social Researcher's Office, 2003; Spencer L, Ritchie J, Lewis J and Dillon L, *Quality in Qualitative Evaluation: A framework for assessing research evidence*, London, Government Chief Social Researcher's Office, 2003; and Pawson R, Boaz A, Grayson L, Long A and Barnes C, *Types and Quality of Knowledge in Social Care: Knowledge review three*, London, Social Care Institute for Excellence, 2003

Table 5: The public value measurement framework

Type of measurement		Stages in the policy cycle				
		Agenda setting	**Task definition**	**Implementation and delivery**	**Outcomes**	**Impact**
Are the methods:	**Appropriate?** • Are they fit for purpose? • Do they meet relevant methodological standards, including method-specific quality standards?					
	Holistic? Do they take into account: • the complexity of the situation? • new public management values such as effectiveness and efficiency? • relevant public value-type values, such as wellbeing, social capital and quality of life?					
	Democratic? Do they allow for: • public involvement? • negotiation between different stakeholders? Are they transparent and accessible? Have they been 'authorised'?					
	Trustworthy? Are the measures and their findings being used appropriately and with integrity?					
	Generating public value? Does the process of measuring create value in and of itself?					

The remaining criteria flow directly from an understanding of public value. Methods should be holistic, take account of the complexity of the situation they are trying to measure, be able to measure values currently associated with the paradigm of new public management such as efficiency and effectiveness, and measure values associated with public value such as wellbeing, social capital and quality of life.

Measures should be democratic, and allow for public involvement and negotiation between different stakeholders. Their process and their findings should be transparent and accessible to the public and specific stakeholders. Measurements should also be endorsed in an authorising environment.

Measures and their findings should be trustworthy, which requires that the criteria relating to appropriateness be met – this might be referred to as 'technical' or 'methodological' trustworthiness. It also requires that methods and their findings are used with integrity, that they are not manipulated in support of one particular argument, taken out of context or generally abused in any manner. Not only must measures be trustworthy, but they must also be seen by the public and all stakeholders to be trustworthy.

An idea unique to the public value approach is that the measuring process in and of itself should actually create public value. Trust in politicians, scientists, public managers and the policy world in general can be created if measurement is appropriate, holistic, democratic and trustworthy. Similarly, democracy may be enhanced if measurement is appropriate, holistic, democratic and trustworthy. Certain methods of measurement, especially those with a participative element, may help to generate values such as human and social capital as participants gain new knowledge or skills from the process, or community cohesion as participants come together to consider issues of importance to their community.

Running across the top of the table from left to right are the stages of the policy process. For simplicity, the stages are represented as operating linearly, although in reality this is far from true (please see Section 2 above for a fuller discussion of the policy cycle). Many research methods and tools for measurement can be applied at several, if not all, stages of the policy process. However, certain methods will be more pertinent to the tasks that need to be performed at various points in the cycle. In the tables that follow we have placed the answers to our criteria questions under the stage of the policy cycle where we feel it is most applicable or most commonly used. Other stages where the method might also be applicable are highlighted too.

As the number of potentially useful methods of measurement is vast, we have selected a sample to apply the framework to of those that we consider to be of particular relevance to public value or that have been discussed in the public value literature. This process has been developmental and the framework has changed somewhat from its first iteration. We hope that others will continue its development through application and that they will find it useful, as we have done, to consider whether or not methods of measurement are appropriate to public value and how to apply them in the most apposite way.

3.3 Measuring what the public values and generating democracy

The public value approach requires the worlds of political and policy decision making to closely involve the public. As Blaug, Horner and Lekhi put it: 'The public value approach is inescapably concerned with the gap between institutions and people.'[32] This gap includes the 'satisfaction gap' (where public satisfaction with public services remains stubbornly low or even falls while government inspections report improvement and the achievement of challenging targets), a lack of engagement with the political process (especially around voter turnout) and a general lack of trust in politicians and public service managers. Measurement can potentially contribute to the closing of this gap, although as is always the case with public value this is not straightforward.

It is clear that we need to establish through some kind of process what it is that the public values and just how much they value it. Central and local government and other institutions have used attitudinal surveys and opinion polls to tap into what the public thinks about anything and everything. Citizens' panels are a more recent evolution of this technique, and one favoured by local authorities.

But while the public value approach is strident on the subject of public involvement in determining broad political agendas and service objectives, aims, and purposes, and maintains that public surveys are one way of doing this, it is also clear that it does not advocate government by opinion poll. Public value recognises that the public will not always value what is necessary for a healthy society or that the public is not always informed enough to realise what is valuable (a phenomenon similar to market failure in public choice theory).

According to Benington, the 'public' is not given but made; it is constantly created and constructed.[33] This construction 'involves a constant battle of ideas and values' as 'competing interests and ideologies' come into play. And the public is not always benign:

[32] Blaug R, Horner L and Lekhi R, *Public Value, Politics and Public Management: A literature review*, London, The Work Foundation, 2006
[33] Benington J, *Creating and Sustaining Public Value in the Health Service*, University of Warwick Health Service Partnership Discussion Paper One, Coventry, Warwick Institute of Governance and Public Management, University of Warwick, 2005

> 'The public realm is under current challenge from tendencies towards, for example, racism, sexism, fascism, fundamentalism, brutalisation, commodification, and consumerism, which fragment the notion of what we have in common as a species and as a public.'[34]

The public needs to be led as well as reacted to, and a good leader will shape public opinion.

This complex understanding of the public has direct implications for the measurement of public value. The idea of public opinion becomes similarly complicated. Should we only be measuring the opinions and attitudes of an informed and competent public? What does a competent public look like? Is it useful to measure uninformed opinion? How should government act in response to a bigoted public? If the public does not agree, how do we measure a variety of opinions and decide which opinions are the most important? Should the measuring process, which is itself charged with generating public value, somehow seek to inform and thus alter opinion? (This last point contradicts much of the underpinnings of science and measurement: that the measurer should observe and not interact with its subject.) How should the results of the measuring process be fed into political debate and policy making? And finally, if we place public opinion at the heart of our political system, and if we decide that what the public values should be measured in ways other than at the ballot box, we should be aware that this potentially constitutes a challenge to the idea of representative democracy. All these questions need to be addressed if we are to be confident that our acts of measurement are to contribute to the production of public value rather than destroy it.

In the face of these challenges we need not throw away our conventional tools of measurement, and nor do advocates of the public value approach ask us to. Instead, we should conceptualise our tools differently and exercise caution in their application. The idea of public value can be used to highlight where and how measuring can be destructive and warn us of the severe consequences of failing in this task. Below we describe a number of methods, both well established and innovative, that are currently used to measure public understanding, attitudes and opinion and begin to use our framework to focus attention on how the techniques might be used to inform a public value approach.

3.4 Measuring public opinion

There are many existing techniques for measuring public understanding, attitudes and opinions. Some, such as deliberative polling, begin to address the issues raised above like that of the importance of an informed public. The Office of the Deputy Prime Minister (ODPM) tells us that English local authorities used 18 different 'participatory methods' in 2001. These are (in order of popularity):

[34] Ibid

- service satisfaction surveys
- complaints/suggestion schemes
- consultation documents
- focus groups
- public meetings
- service user forums
- citizens' panels
- area/neighbourhood forums
- community plans/needs analysis
- other opinion polls
- question and answer sessions
- co-option/committee work
- issue forums
- shared interest forums
- visioning exercises
- user management of services
- referendums
- citizens' juries.[35]

And as Blaug, Horner and Lekhi state, still more methods are to be found being used in the fields of planning, international development and corporate team building. Some of these are more closely associated with the term 'measurement' than others: while we are familiar with the idea of the public meeting as an opportunity for 'ordinary people' to get their voice heard, we more readily associate satisfaction surveys with actually measuring public opinion.

Each method has methodological strengths and weakness (there are surely even 'right ways' and 'wrong ways' of running a public meeting) and are subject to criticism both from the public value debate and independently of it. Thus we don't need public value to tell us that our measurements should be accurate and carried out to industry-agreed standards, but it does remind us of the implications for values like trust and democracy if they aren't and adds an extra urgency to this concern. Thus many of the important criticisms of measurement methods from the public value camp, such as issues around research fatigue and measurement being 'quick and dirty'[36], can be answered by an appeal to proper measurement that is carried out with integrity as opposed to sloppily or with an agenda. Similarly, many of the criticisms focused around trust in measurements can be addressed by considering how the findings of the measurement process are actually used: survey results will not add value if they gather dust in a local authority's cupboard instead of feeding in to the policy making process.

[35] Office of the Deputy Prime Minister, *Public Participation in Local Government: A survey of local authorities*, London, ODPM, 2002

[36] Blaug R, Horner L and Lekhi R, *Public Value, Politics and Public Management: A literature review*, London, The Work Foundation, 2006

(Alarmingly, this is indeed the fate of many specially commissioned surveys.)[37] The public value critique that government-favoured methods of consultation are overwhelmingly top-down serves as a reminder that appropriate methods must be selected. For instance, top-down consultation probably is not the best way to achieve active public participation in decision making. It may well be that methods more appropriate to the public value paradigm need to be developed, but for now we do have a vast array of methods that we can, tentatively, begin to employ with a public value bent.

3.4.1 Consensus conferences

A consensus conference is designed to incorporate the perspectives of the lay public in the assessment of new scientific and technological developments.[38] A panel of around six to ten laypeople questions experts about a controversial scientific or technological subject, assesses the experts' responses, reaches a consensus about the subject and reports its conclusions. The panel is selected from volunteers who respond to an advert to participate. The first stage of recruiting the panel is therefore self-selecting on the part of the participants. In the second stage recruitment panel members are chosen on the basis of a number of socio-demographic characteristics such as age, gender, education, occupation and area of residence. The aim is to produce a panel of people with diverse perspectives because it is not possible for such a small group to be statistically representative of the public. How the 'experts' are selected is potentially contentious, as the panel overall may exclude key areas of evidence or present evidence of unequal rigour and standing, which laypeople are not in a position to judge.

Consensus conferences have been used in Denmark since the 1980s. The Danish Board of Technology regularly runs conferences addressing contentious scientific or technological issues. Although the conference panel has no statutory authority, they are potentially influential. For example, the panel's recommendations on genetic engineering in industry and agriculture led to the exclusion of transgenic animals from the first governmental biotechnology research and development programme. Such an impact is facilitated by the well-established relationship that the Danish Board of Technology has with the Danish parliament through its nine-member Research Committee. Conferences are also held in the parliament building at times when topics for discussion are already high on the agenda. Media attention is strong and this drives public debate more widely, thus maximising public engagement. It is possible that such conferences contribute

---

[37] Kelly G, Muers S and Mulgan G, *Creating Public Value: An analytical framework for public service reform*, London, Cabinet Office, 2002. Aside from a vague duty for findings not to be blatantly manipulated and done so while still bearing the name of independent 'measurers', the people doing the measuring have usually bowed out of the process at the point where findings are handed over to the client. It could be argued that the public value position encourages the continuing involvement of independent scientists throughout the policymaking process

[38] Scottish Office, *Involving Civil Society in the Work of Parliaments*, Scottish Office. Available at www.scotland.gov.uk/government/devolution/cpsp-00.asp

to the high levels of confidence that the Danish public has in the institutions responsible for scientific and technological development.

A single consensus conference has so far been convened in the UK. It was held in 1994 on the subject of plant biotechnology in food and agriculture. Although around 300 people attended the conference daily, the level of media coverage was disappointing and the chair of the Lords Select Committee on Biotechnology told the conference that parliament would not be influenced by the lay panel's report in any significant way.

Table 6: Consensus conferences and the measurement framework

Consensus conferences		Stages in the policy cycle				
		Agenda setting	**Task definition**	**Implementation and delivery**	**Outcomes**	**Impact**
Are consensus conferences:	**Appropriate?** • Are they fit for purpose? • Do they meet relevant methodological standards, including method-specific quality standards?		• Should meet relevant methodological standards			
	Holistic? Do they take into account: • the complexity of the situation? • new public management values such as effectiveness and efficiency? • relevant public value-type values, such as wellbeing, social capital and quality of life?		• The information-rich and deliberative process allows for complex issues to be addressed			
	Democratic? Do they allow for: • public involvement? • negotiation between different stakeholders? Are they transparent and accessible? Have they been 'authorised'?		• Involves a small number of members of the public. The panel will produce a set of recommendations or a report, which may or may not be considered by decision makers • The process informs the panel members. The wider public may be involved and informed through media coverage and access to the process. An informed public is necessary for the creation of democracy in the public value approach • Panel members reach a consensus • Democracy will be enhanced if results are used appropriately to inform the political process			
	Trustworthy? Are the measures and their findings being used appropriately and with integrity?		• Trust in the process can be enhanced by employing an independent organisation to act as facilitator • Transparency of method and how findings are used will enhance trustworthiness			
	Generating public value? Does the process of measuring create value in and of itself?		• Generates democracy (see above) • Has the potential to generate trust and empower participants through increased knowledge			

3.4.2 Citizens' juries

Citizens' juries have been experimented with in the UK since the 1990s and have been used in the United States, Germany, Spain, The Netherlands and Palestine. Citizens are chosen, often at random, to form a jury that hears information from and directs questions to witnesses about an issue in order to inform the setting of a policy agenda or to choose between a number of policy options. The jury will deliberate and then deliver its verdict in the form of a set of recommendations or report. The body sponsoring the jury, which will be a government department, a local authority or other agency, usually undertakes to respond in some way to the jury's findings. Citizen's juries have been held on subjects including local planning, energy, technology and communication, the environment and transport.[39]

Numbers of jurors usually range from 12 to 16 people who are chosen to participate in the process using either random or stratified sampling. The pioneer of citizens' juries in Germany advocates random sampling of participants, arguing that it is important for all citizens to have an equal chance of being selected. However, the UK and US have tended to stratify according to demographics. German juries are usually larger, consisting of around 25 members of the public, and several juries are run either concurrently and/or in parallel: the largest project in Germany involved some 500 citizens. The acceptance rate (which could be seen as the equivalent of the response rate in social surveys or opinion polling) for invitations to participate in juries in Germany runs at about 15-20 per cent for issues involving technological questions and 25-40 per cent for local planning issues. Perhaps unsurprisingly, when the public is interested in the topic for consideration they are more interested in taking part in the process. Participants often report experiencing a sense of empowerment and many become more active citizens after being part of a jury.

In the UK, proponents of the use of citizens' juries include the Institute of Public Policy Research (IPPR), the King's Fund Policy Institute and the Local Government Management Board (LGMB). Each has sponsored a series of pilot projects, normally in collaboration with health authorities or local government. The pilots all used an independent facilitating organisation to ensure that the process was fair. There is concern that this good practice is not always reproduced as more organisations have begun to make use of citizens' juries. There has also been debate surrounding the representativeness of juries and claims that they merely give the 'usual suspects' (those members of the public who would normally make themselves heard) another voice in the political process.

[39] Ibid

Table 7: Citizens' juries and the measurement framework

Citizens' juries		Stages in the policy cycle				
		Agenda setting	**Task definition**	**Implementation and delivery**	**Outcomes**	**Impact**
Are citizens' juries:	**Appropriate?** • Are they fit for purpose? • Do they meet relevant methodological standards, including method-specific quality standards?		• Should meet relevant methodological standards			
	Holistic? Do they take into account: • the complexity of the situation? • new public management values such as effectiveness and efficiency? • relevant public value-type values, such as wellbeing, social capital and quality of life?		• The information-rich and deliberative process allows for complex issues to be addressed			
	Democratic? Do they allow for: • public involvement? • negotiation between different stakeholders? Are they transparent and accessible? Have they been 'authorised'?		• Involves a small number of members of the public. The jury will produce a set of recommendations or report, which the sponsoring body usually undertakes to respond to • The process informs the jurors. The wider public may be involved and informed through media coverage and access to the process. An informed public is necessary for the creation of democracy in the public value approach • Jurors reach a consensus • There is some concern that juries give voice to 'the usual suspects' as the rate of positive response to an invitation to participate is often low • Democracy will be enhanced if results are used appropriately to inform the political process			
	Trustworthy? Are the measures and their findings being used appropriately and with integrity?		• Trust in the process can be enhanced by employing an independent organisation to act as facilitator • Transparency of method and how findings are used will enhance trustworthiness			
	Generating public value? Does the process of measuring create value in and of itself?		• Generates democracy (see above) • Has the potential to generate trust and empower participants through increased knowledge			

3.4.3 Attitudinal surveys and opinion polling

Parsons describes public opinion as being 'to the political market what consumer demand is to the economic market place'.[40] Some would go so far as to endorse 'government by opinion poll' claiming that politicians would no longer claim a political mandate, but instead would know what the public felt about a particular issue. Gallup, the pioneer of modern-day opinion polling, believed in learning and reporting 'the will of the people'.[41] Following Gallup, we can immediately see the method's potential to contribute to the measurement of public value, although the issues raised above must weigh heavily on our minds. The use of opinion polls is certainly on the increase, apparently in proportion to the public's increasing lack of enthusiasm for the ballot box. Pratchett, director of the local governance research unit at De Montfort University, states that opinion polling is now an everyday part of local authorities' service delivery[42] and MORI claims to have carried out in excess of 1,000 projects for 250 local authorities over the last decade.[43]

While survey and polling techniques have become highly sophisticated and the methodological literature on sampling, eliminating bias and questionnaire design is well developed, measuring people's attitudes towards, opinions on and general satisfaction with public services remains tricky. Levels of satisfaction with services vary systematically between users and non-users and respondents are more likely to be happy with the running of their authority if the party they voted for is in control. Moreover, issues are more salient to some people than others and this tends to affect their satisfaction threshold; respondents will happily give opinions on non-existent issues and people will give different answers to the same questions if asked in a different order. (This phenomenon was beautifully illustrated in an episode of the television series *Yes, Prime Minister*, the transcript of which is reproduced for your amusement – and to demonstrate a serious technical point – in Box 1 overleaf.)

A well-established and respected example of an attitudinal survey is that of the British Social Attitudes Survey.[44] Undertaken annually, the survey asks a randomly selected sample of some 3,300 adults their views on a wide range of topics including housing, work, transport, government spending, voting habits, religion, racism and illegal drugs. Questions are repeated periodically in order to generate longitudinal data, and new questions are added to reflect current agendas and debates. Questions are developed in collaboration with grant-giving bodies and government departments, but the survey is independent and a condition of its core funding is that the survey series' researchers control all aspects of question design and data analysis. NatCen attribute the survey's success and longevity partly to this independent status.

---

[40] Parsons W, *Public Policy: An introduction to the theory and practice of policy analysis*, Edward Elgar, Cheltenham, 1985

[41] See the Gallup Organisation's history of George Gallup at http://www.gallup.com/content/?ci=21364

[42] Appleton J, *Government by Opinion Poll*, Spiked Politics, 2005. Available at www.spiked-online.com/Articles/0000000CA99A.htm

[43] See http://www.mori.com/localgov/index.php

[44] National Centre for Social Research, *Social Attitudes and Elections* (undated). Web publication available at www.natcen.ac.uk/natcen/pages/or_socialattitudes.htm

Box 1: How questionnaire design can influence response

Sir Humphrey: 'You know what happens: nice young lady comes up to you. Obviously you want to create a good impression, you don't want to look a fool, do you? So she starts asking you some questions: Mr Woolley, are you worried about the number of young people without jobs?'

Bernard Woolley: 'Yes'

Sir Humphrey: 'Are you worried about the rise in crime among teenagers?'

Bernard Woolley: 'Yes'

Sir Humphrey: 'Do you think there is a lack of discipline in our comprehensive schools?'

Bernard Woolley: 'Yes'

Sir Humphrey: 'Do you think young people welcome some authority and leadership in their lives?'

Bernard Woolley: 'Yes'

Sir Humphrey: 'Do you think they respond to a challenge?'

Bernard Woolley: 'Yes'

Sir Humphrey: 'Would you be in favour of reintroducing National Service?'

Bernard Woolley: 'Oh...well, I suppose I might be.'

Sir Humphrey: 'Yes or no?'

Bernard Woolley: 'Yes'

Sir Humphrey: 'Of course you would, Bernard. After all you told me, you can't say no to that. So they don't mention the first five questions and they publish the last one.'

Bernard Woolley: 'Is that really what they do?'

Sir Humphrey: 'Well, not the reputable ones no, but there aren't many of those. So alternatively the young lady can get the opposite result.'

Bernard Woolley: 'How?'

Sir Humphrey: 'Mr Woolley, are you worried about the danger of war?'

Bernard Woolley: 'Yes'

Sir Humphrey: 'Are you worried about the growth of armaments?'

Bernard Woolley: 'Yes'

Sir Humphrey: 'Do you think there is a danger in giving young people guns and teaching them how to kill?'

Bernard Woolley: 'Yes'

Sir Humphrey: 'Do you think it is wrong to force people to take up arms against their will?'

Bernard Woolley: 'Yes'

Sir Humphrey: 'Would you oppose the reintroduction of National Service?'

Bernard Woolley: 'Yes'

Sir Humphrey: 'There you are, you see Bernard. The perfect balanced sample.'

Excerpt from *Yes, Prime Minister*, series 1, episode 2, 'The Ministerial Broadcast'.
First aired on the BBC, 16 January 1986.

Table 8: Attitudinal surveys and opinion polling, and the measurement framework

Attitudinal surveys and opinion polling		Stages in the policy cycle				
		Agenda setting	**Task definition**	**Implementation and delivery**	**Outcomes**	**Impact**
Are attitudinal surveys and opinion polling:	**Appropriate?** • Are they fit for purpose? • Do they meet relevant methodological standards, including method-specific quality standards?			• Must meet relevant methodological standards		
	Holistic? Do they take into account: • the complexity of the situation? • new public management values such as effectiveness and efficiency? • relevant public value-type values, such as wellbeing, social capital and quality of life?			• The technique may struggle to capture complexity as it simply provides a 'snapshot' of the potentially uninformed and unconsidered opinions of individuals that constitute the 'public' • Traditionally, surveys and polls have found it difficult to capture public value-type values, although more effort has recently been focused in this area		
	Democratic? Do they allow for: • public involvement? • negotiation between different stakeholders? Are they transparent and accessible? Have they been 'authorised'?			• Involve the public, but participation is passive • Can represent a range of views (consensus is not an aim of the exercise) • Democracy will be enhanced if results are used appropriately to inform the political process		
	Trustworthy? Are the measures and their findings being used appropriately and with integrity?			• Trust in the process can be enhanced by employing an independent organisation to undertake the survey • Transparency of method and how findings are used will enhance trustworthiness		
	Generating public value? Does the process of measuring create value in and of itself?			• If the public feels that their views are important, value can be created • Surveys that use and train interviewers from the population being surveyed generates human capital and also has the potential to generate social capital		

3.4.4 Deliberative opinion polling

Developed by American political scientist James Fishkin, deliberative opinion polling explores public opinion on a subject both before and after people are given time to think about the subject in question. Unlike standard opinion polling, which takes a snapshot of opinions held by an often uninformed public, the process of deliberative polling both surveys opinion and provides the opportunity for participants to become informed about the survey subject. Thus deliberative polling can be: descriptive, in that it plots considered public opinion against first reactions; predictive, in that it shows how public opinion might change if an issue were to become more salient; and prescriptive, in that it points to what an 'informed and reflective citizenry' might wish policymakers to do.[45]

A statistically representative sample of between 250 and 600 members of the public is chosen to participate in the poll. Participants are polled to record their 'prior' views on the subject. Participants are then provided with briefing information about the subject of the poll detailing the main points of contention. Participants then attend an event of between 2 and 4 days where they are 'immersed' in the issues through focus groups and plenary sessions. Here they can question experts and politicians who defend a variety of different positions and deliberate among themselves. The participants are then re-polled using the same questionnaire as that given to them prior to the event.[46]

A number of deliberative opinion polls have been run in the UK sponsored by Channel 4. The first, in 1994, was on the subject 'Rising Crime: What can we do about it?' Other polls have covered the future of the UK in Europe, the future of the monarchy, and voting intentions for the general election. Deliberative polls have also been used in the US. Unlike citizens' juries and consensus conferences, deliberative polling doesn't aim to produce a consensus among participants and is therefore a more appropriate measure if the diversity of public opinion is of concern.

---

[45] Ibid

[46] Scottish Office, *Involving Civil Society in the Work of Parliaments*, Scottish Office. Available at www.scotland.gov.uk/government/devolution/cpsp-00.asp

Table 9: Deliberative opinion polling and the measurement framework

Deliberative opinion polling		Stages in the policy cycle				
		Agenda setting	**Task definition**	**Implementation and delivery**	**Outcomes**	**Impact**
Is deliberative opinion polling:	**Appropriate?** • Is it fit for purpose? • Does it meet relevant methodological standards, including method-specific quality standards?		• Meets relevant methodological standards			
	Holistic? Does it take into account: • the complexity of the situation? • new public management values such as effectiveness and efficiency? • relevant public value-type values, such as wellbeing, social capital and quality of life?		• Can measure informed opinion on complex subjects			
	Democratic? Does it allow for: • public involvement? • negotiation between different stakeholders? Is it transparent and accessible? Is it 'authorised'?		• Involves the public. The process is quite passive, but less so than standard opinion polling • The process actually informs the public. An informed public is necessary for the creation of democracy in the public value approach • Can represent a range of views (consensus is not an aim of the exercise) • Democracy will be enhanced if results are used appropriately to inform the political process			
	Trustworthy? Is the measure and its findings being used appropriately and with integrity?		• Trust in the process can be enhanced by employing an independent organisation to act as facilitator • Method transparency and how findings are used will enhance trustworthiness			
	Generating public value? Does the process of measuring create value in and of itself?		• Generates democracy (see above) • Has the potential to generate trust and empower participants through increased knowledge			

3.4.5 Citizens' panels

Citizens' panels involve between 500 and 3,000 members of the public. They tend to be used by local authorities to gauge local opinion on a variety of issues around policy options and the delivery of services. Panel organisers attempt to achieve a statistically representative sample of the area's population, which is then weighted for gender, age, ethnic background and occupation. Panel members may be surveyed as little as once a year or may be consulted more frequently. Panel members usually undertake to serve for a period of time of about 3 years, after which new members are recruited. This is largely to avoid panel members becoming unrepresentative of the local population; it is suspected that a significant number of panel members become more informed on the survey subjects as the process of regularly being asked questions about them arouses interest or highlights knowledge gaps. This phenomenon, which is considered a methodology issue for citizens' panels, may actually serve to generate public value in the form of contributing to a more informed and engaged citizenry. In the UK, citizens' panels were first used by the City of Bradford Metropolitan District Council, the London Borough of Lewisham and Southampton City Council[47], and were used by 71 per cent of English local authorities in 2001.[48]

[47] Ibid

[48] Office of the Deputy Prime Minister, *Public Participation in Local Government: A survey of local authorities*, London, ODPM, 2002

Table 10: Citizens' panels and the measurement framework

Citizens' panels		Stages in the policy cycle				
		Agenda setting	**Task definition**	**Implementation and delivery**	**Outcomes**	**Impact**
Are citizens' panels:	**Appropriate?** • Are they fit for purpose? • Do they meet relevant methodological standards, including method-specific quality standards?		• Meets relevant methodological standards			
	Holistic? Do they take into account: • the complexity of the situation? • new public management values such as effectiveness and efficiency? • relevant public value-type values, such as wellbeing, social capital and quality of life?		• The technique may struggle to capture complexity as it simply provides a 'snapshot' of the potentially uninformed and unconsidered opinions of individuals who constitute the 'public' • Traditionally, surveys and polls have found it difficult to capture public value-type values, although more effort has recently been focused on this area			
	Democratic? Do they allow for: • public involvement? • negotiation between different stakeholders? Are they transparent and accessible? Have they been 'authorised'?		• Involves the public, but participation is passive • Can represent a range of views (consensus is not an aim of the exercise) • Democracy will be enhanced if results are used appropriately to inform the political process			
	Trustworthy? Are the measures and their findings being used appropriately and with integrity?		• Trust in the process can be enhanced by employing an independent organisation to undertake the survey • Transparency of method and how findings are used will enhance trustworthiness			
	Generating public value? Does the process of measuring create value in and of itself?		• In that the public feels its views are important, value can be created			

3.4.6 Cost-benefit analysis and contingent valuation

A method commonly used by government for evaluating policy, both ex ante and ex post, is cost-benefit analysis (CBA). Although not required by the Treasury as a justification for policy decisions, the method features heavily in the *Green Book* and pressure is felt across government and the evaluation community to employ 'hard' techniques that assign monetary values to policy outcomes. Put simply, CBA assigns a monetary value to all the costs and all the benefits associated with a policy or programme and adds them up. If the answer is positive, then the benefits outweigh the costs and the policy is judged to be worthwhile. If the answer is negative, then the costs are too high. The difficulty of the technique, and one of particular concern to public value, is that intangibles resist monetisation.

For non-marketable items – those that cannot be bought or sold – two methods of valuation can be used. Revealed preferences use actual transactions such as the amount of money people spend travelling to a place of interest as a proxy measure or estimate of their value. Stated preferences rely on hypothetical behaviour and willingness to pay in order to estimate value. The 'hard' numerical inputs and outputs of a CBA can encourage one to forget that they are tools for estimating that require judgement in their application. But a good CBA will make uncertainties explicit and express the results as a range of values rather than as a single-point estimate.[49]

Used widely in environmental economics, contingent valuation is a method of measuring stated preferences. It can be used to estimate both use and non-use values.[50] A sample of the public, ideally one that is statistically representative, is surveyed to see how much they would be willing to pay for something or how much compensation they would want in order to give something up. Question design is tricky and informed by initial interviews and/or focus groups around the subject under consideration. The method gets its name as respondents are asked to state their willingness to pay contingent on a specific hypothetical scenario and description of a service. Because contingent valuation relies on hypothetical behaviour it can estimate non-use, but this source of strength is also its greatest source of weakness: hypothetical and actual behaviour may differ. The conceptual, empirical, and practical problems associated with the method are hotly debated in the economics literature.[51]

---

[49] ESRC Global Environmental Change Programme, 'Making Environmental Decisions: Cost-Benefit Analysis, Contingent Valuation and Alternatives', *Proceedings from the Centre for the Study of Environmental Change/Green Alliance Practitioners' Seminar,* held on 23 January 1997 at McKenna & Co, London. Available at: www.sussex.ac.uk/Units/gec/pubs/reps/gralconf.htm

[50] Non-use values are where *non*-users derive satisfaction from something's mere existence. For example, I derive great satisfaction from knowing that the rain forests exist without ever being likely to see them, and their value to me is not reliant on my 'use' of them.

[51] Ecosystem Valuation, *Contingent Valuation Method*. Web publication available at www.ecosystemvaluation.org/contingent_valuation.htm

Contingent valuation has been used in the US for environmental decision making. The State of California commissioned a contingent valuation to form part of an environmental impact report. A telephone survey was conducted with respondents who had been provided in advance with information booklets. Respondents were asked how they would vote in a hypothetical referendum. While it cannot be claimed that the survey results were a deciding factor, the California Water Resources Control Board opted for the policy that respondents valued the most. In the UK, the British Library has recently employed contingent valuation in order to value the contribution of the library to society. Over 2,000 people were interviewed, representing various types of user and non-user. The British Library holds that the method is appropriate given that the value it adds takes many intangible forms – economic, cultural, social and intellectual – and is of value to both users and non-users.[52]

[52] British Library, *Measuring Our Value: Results of an independent economic impact study commissioned by the British Library to measure the Library's direct and indirect value to the UK economy,* London, British Library, 2004. Available at: www.bl.uk/pdf/measuring.pdf

Table 11: Contingent valuation and the measurement framework

Contingent valuation		Stages in the policy cycle				
		Agenda setting	**Task definition**	**Implementation and delivery**	**Outcomes**	**Impact**
Is contingent valuation:	**Appropriate?** • Is it fit for purpose? • Does it meet relevant methodological standards, including method-specific quality standards?				• Should meet relevant methodological standards	
	Holistic? Does it take into account: • the complexity of the situation? • new public management values such as effectiveness and efficiency? • relevant public value-type values, such as wellbeing, social capital and quality of life?				• Contingent valuation seeks to measure intangibles, thus ensuring that they are valued in the decision-making process • Can be organised so as to inform respondents about the issues before they are surveyed, thus enabling fairly complex issues to be addressed	
	Democratic? Does it allow for: • public involvement? • negotiation between different stakeholders? Is it transparent and accessible? Has it been 'authorised'?				• Involves the public, but the method is 'top-down' and passive • Can represent a variety of views (consensus is not an aim of the exercise)	
	Trustworthy? Is the measure and its findings being used appropriately and with integrity?				• Trust in the process can be enhanced by employing an independent organisation to act as facilitator • Transparency of method and how findings are used will enhance trustworthiness	
	Generating public value? Does the process of measuring create value in and of itself?				• Guards against the destruction of public value in that it recognises the importance of values that resist measurement	

3.4.7 Performance measurement and public value

Performance measurement is more helpfully conceived of as a family of measures rather than a single method. Used to inform decision making at most points in the policy cycle, it is most strongly associated with the achievement of targets or the introduction of performance indicators that relate to the processes of implementation, policy outputs and, increasingly, policy outcomes. Performance measurement is a child of new public management. Its roots are in the Thatcher governments of the 1980s when it was introduced as a means of top-down control, part of a suite of measures designed to bring the discipline of the market into the public sector. The era of performance measurement has continued under the Blair government, with league tables in education, star ratings in health, Comprehensive Performance Assessment and Best Value Performance Indicators in local government, and other performance measurement regimes becoming so much a part of public service management that critics talk of a pervading audit culture.

The rationale behind the use of performance measures is that of measurement in general. As Osborne and Gaebler famously argued: if you don't measure results, you can't tell success from failure; if you can't see success you can't reward it; if you can't reward success, you're probably rewarding failure; if you can't see success, you can't learn from it; if you can't recognise failure, you can't correct it; if you can demonstrate results, you can win public support; what gets measured gets done.[53] If frequency of use is an indication of utility, then performance measures must be useful indeed. But some of performance measurement's failings, so-called 'perverse effects', have been hitting the headlines of late. For example, consider the pre-election revelation that in response to the target that no patient should wait more than 48 hours to see a GP, some surgeries refused to book appointments more than 48 hours in advance.

These phenomena, while being discussed independently of the public value paradigm, are highlighted by the public value approach. It reminds us to ask 'what is measurement for?' and 'are we measuring the right things in the right way?' And it begins to answer these questions. For instance, we certainly should be paying great attention to intangibles. Just because they are difficult to measure does not mean that they don't count. Attempts to improve performance measurement are ongoing and effort is being made to measure intangibles and public value-type values. The Audit Commission has now developed quality-of-life themes as part of its Area Profiles model.[54] And those in government who are charged with the task of setting targets are aware of the need for joined-up thinking so that targets are mutually reinforcing rather than being in competition with each other, and are making targets 'smart' to avoid perversity. Despite its vocal criticism, the public

[53] Cited in Blaug R, Horner L and Lekhi R, *Public Value, Politics and Public Management: A literature review,* London, The Work Foundation, 2006

[54] Audit Commission, *Area Profiles Phase II: Quality of life themes*, Audit Commission, 2005. Available at www.areaprofiles.auditcommission.gov.uk/InformationPages/qualityoflife.pdf

value approach does not necessarily call for the trashing of performance measurement, but demands its careful and considered application as it does with all existing measures.

Table 12: Performance measurement and the measurement framework

Performance measurement		Stages in the policy cycle				
		Agenda setting	**Task definition**	**Implementation and delivery**	**Outcomes**	**Impact**
Are methods of performance measurement:	**Appropriate?** • Are they fit for purpose? • Do they meet relevant methodological standards, including method-specific quality standards?				• Systems of performance measurement are particularly prone to 'perverse effects'. As the government/non-profit sectors become more familiar with the technique, learning to avoid inappropriate measures is being applied	
	Holistic? Do they take into account: • the complexity of the situation? • new public management values such as effectiveness and efficiency? • relevant public value-type values, such as wellbeing, social capital and quality of life?				• Performance measurement more readily lends itself to measuring effectiveness and efficiency than other values, as it springs from the new public management paradigm • How to use performance measurement to track values such as social capital and wellbeing is starting to be considered	
	Democratic? Do they allow for: • public involvement? • negotiation between different stakeholders? Are they transparent and accessible? Have they been 'authorised'?				• Traditionally a 'top-down' approach, the citizen could exercise influence by assuming the role of 'customer' of the services ultimately delivered • Experiments to involve the public in the setting of targets and objectives are underway	
	Trustworthy? Are the measures and their findings being used appropriately and with integrity?				• Performance measures can encourage managers to deliberately manipulate measurement information (a particular type of 'perverse effect') • The public may not trust the findings of performance measurement if they do not ring true with their personal experience	
	Generating public value? Does the process of measuring create value in and of itself?				• 'Perverse effects' of measurement can destroy value	

3.4.8 The balanced scorecard and the public value scorecard

Mark Moore developed the public value scorecard in response to what he saw as the failings of the widely-used balanced scorecard when applied to the not-for-profit sector.[55] Originally designed for the private sector by Robert Kaplan, the balanced scorecard has been embraced by not-for-profit and voluntary organisations as it is a tool for strategy and performance management that emphasises the importance not only of measuring the financial bottom-line, but also of non-financial measures.

While Moore is an admirer of Kaplan's understanding that the financial information surrounding an organisation is not sufficient for informing a successful business strategy, Moore argues that the primary concern of the balanced scorecard remains one of maximising organisational profit – a purpose, by definition, not found in the not-for-profit sector. Thus the four areas where the balanced scorecard requires measurements (the financial perspective, the customer perspective, the operational perspective, and the learning and growth perspective) are not appropriate for public and voluntary organisations, or at the very least require a different insight in order to be applicable to them.

In terms of the financial perspective, Moore argues that while financial efficiency and effectiveness should always be a concern to any organisation regardless of its purpose, it should not be given pride of place as the balanced scorecard gives it in a tool for strategic planning and measurement used by not-for-profit organisations. In this context, making best use of an organisation's resources will not be about making money, but will instead be subsumed in the organisation's charitable or social mission. As Moore puts it: 'Their value is not measured primarily by the willingness of customers to plunk down money to consume the goods and services they offered.'[56]

From the customer perspective, as Moore's quote above highlights, the notion of a 'customer' is somewhat complicated in the not-for-profit sector. While it is possible to conceive of the end recipient benefiting from an organisation's work as its customer, they are usually not in a position to pay for that organisation's services – often the raison d'être for the organisation in the first place. There are also people further 'upstream' who might be seen to act as 'customers' and whose wishes need to be considered. Benefactors, government and the wider public can also be thought of as 'customers'.

From an operational point of view, not-for-profit organisations function in an atmosphere of partnership rather than competition. Organisations may compete for funding, but unlike profit-making firms they do not seek to capture and hold

[55] Moore M H, *The Public Value Scorecard: A rejoinder and an alternative to 'Strategic Performance Measurement and Management in Non-Profit Organizations' by Robert Kaplan*, HCNO Working Paper Series, May 2002

[56] Ibid

on to a significant share of the market. Rather, their goal should be to strengthen the sector as a whole through co-operation and knowledge sharing.

In the public value scorecard, Moore attempts to provide a strategy and performance measurement tool applicable to these unique circumstances. He asserts that not-for-profit organisations should focus on the three inter-related areas of value or social mission, legitimacy and support, and operational capacity. This focus takes account of the issues discussed above and directs measurement to support strategy in a manner appropriate to the voluntary sector.

The public value scorecard functions very much like strategic and performance management tools designed in ignorance of a public value approach. It functions in accordance with the usual logic that organisational mission defines higher level objectives, which in turn define lower level objectives that are intended to bring about the desired outputs, outcomes and impacts, all of which need to be measured in order to gauge progress. What the public value scorecard does is recognise the potential perverse effects or inadequacies of certain systems of performance management and offer a coherent framework for guiding measurement away from some of these pitfalls. The lack of any real novelty in terms of measurement techniques accompanying the public value scorecard again signals that the literature on measurement in the public value approach is in its infancy, and prompts us to ponder whether this new paradigm needs new tools for measurement or if our existing techniques can be adequately adapted.

Table 13: The public value scorecard and the measurement framework

Public value scorecard		Stages in the policy cycle				
		Agenda setting	**Task definition**	**Implementation and delivery**	**Outcomes**	**Impact**
Is the public value scorecard:	**Appropriate?** • Is it fit for purpose? • Does it meet relevant methodological standards, including method-specific quality standards?			• The public value scorecard was designed in order to avoid an inappropriate focus and thus 'perverse effects'. As with all performance measurement systems, 'perversity' will remain a concern		
	Holistic? Does it take into account: • the complexity of the situation? • new public management values such as effectiveness and efficiency? • relevant public value-type values, such as wellbeing, social capital and quality of life?			• The public value scorecard is designed to take into account the particular context surrounding not-for-profit organisations • It is a tool for performance management and is therefore concerned with effectiveness and efficiency while being sensitive to the fact that these are not the ultimate aim, but factors to support the organisation's social mission		
	Democratic? Does it allow for: • public involvement? • negotiation between different stakeholders? Are they transparent and accessible? Have they been 'authorised'?			• The public value scorecard emphasises the importance of 'legitimacy and support', which directs attention to benefactors, government, beneficiaries, and the wider public		
	Trustworthy? Are the measures and their findings being used appropriately and with integrity?			• An insistence on transparency and a consideration of 'legitimacy and support' will help to make the method trustworthy		
	Generating public value? Does the process of measuring create value in and of itself?			• A focus on social mission should ensure that measurement does not destroy value		

3.4.9 Measuring outcomes and impacts

Because of the emphasis on performance measurement encouraged by new public management practices and by the setting of targets and public service agreements there has been a great deal published over the last ten years in relation to the measurement of outcomes and impacts of public services.[57]

Many of these have a primary focus on the assessment of the specific outcomes from policies or programmes, although there has also been the development of tools for the assessment of broad impacts from a raft of policy interventions, such as the Best Value regimes or Public Service Agreement (PSA) targets. In most cases, this has often involved the identification of proxy indicators for some broad aspects of impact such as general wellbeing or improvements in quality of life. For example, economic indicators such as economic growth have frequently been used as an indicator for general wellbeing, particularly in developing countries, although this has increasingly been criticised as too narrow an indicator.[58]

Following the earlier arguments, the implications of applying the notion of public value to the measurement of outcomes and impacts suggests two things: that the outcomes and impacts themselves may be conceived in broader and more holistic terms, and that the processes through which any measures are developed and used need to be considered carefully.

In terms of broader outcomes, in Section 2 we identified a number of clusters of values that relate to the outcome of public services, which go beyond the delivery of specific service outcomes. These include:

- quality of life, wellbeing and happiness
- social capital/social cohesion/social inclusion
- safety and security
- equality/tackling deprivation and social exclusion
- promoting democracy and civic engagement.

Each of these areas has been subject to considerable research and discussion, and it would be difficult to do justice to this in the present document. However, below we provide a brief illustration of some of the central challenges in relation to the quality of life, wellbeing and happiness clusters. This is a useful one to start with, particularly as there is considerable overlap between this and many of the other dimensions – see for example under the section on the breadth of the concept.

---

[57] Including the Cabinet Office's *Regulatory Impact Assessment Guidance* (2005), the Strategy Unit's *Impact Assessment and Appraisal Guidance Checklist of Policymakers* (2002) and others mentioned earlier such as the *Green Book Guidance on Policy Appraisal* and the *Magenta Book: Guidance Notes on Policy Evaluation.*

[58] Layard R, *Happiness: Lessons from a new science*, London, Allen Lane, 2005

3.4.9.1 Quality of life, wellbeing and happiness

These three concepts represent related, and to some extent, overlapping dimensions of public value and have each been subject to considerable debate and discussion, and the development of 'measures'. All three pose similar challenges in relation to the fact that they:

- have a substantial subjective element
- are very broad concepts that can be operationalised very differently in different contexts
- have both an individual and collective dimension.

Each of these concepts involve a subjective assessment by the individual that relates to their particular set of expectations and values, and which vary considerably from person to person. Initially, this has led to some reluctance to attempt to develop an objective scale as noted in recent literature related to the operationalisation of concepts such as 'happiness'.[59] However, in the health field criticisms that an overt focus on the clinical outcomes of intervention has stimulated a strong tradition of developing scales to capture subjective experiences of 'wellbeing' and 'quality of life', such as QLYS, SF36 and the Rosser scales. These attempt to combine a number of different aspects identified initially by a group of patients into a composite scale of patient experience and functioning. However, one difficulty in using a composite scale of this kind across a wide range of different health conditions is that the range of aspects included is always limited, which limits its application in particular situations. For example, Euroquol is a widely used quality-of-life scale that excludes the dimension of sleep disturbance. This limits its ability to capture change in any conditions where this is a major symptom.

There have also been difficulties in incorporating such scales into policymaking or policy assessment rather than using them for the purpose for which they were originally developed, ie the assessment of outcomes from specific clinical intervention. This was highlighted in the USA by the Oregon project, in which quality-of-life scales were paired with cost-effectiveness indicators in an attempt to consult the public about the prioritisation of conditions/treatments for funding under the Medicare budget. This initially returned some strange results (eg appendectomies were given lower priority than the treatment of tooth decay) and the resulting prioritisation failed to reflect regional and cultural diversity. The complexity of making adjustments to meet these concerns meant that the final results were determined more by health commissioners than by the public. The use of quality-of-life scales was also challenged by disability groups and was finally excluded from the equation.

For individual versus collective measures, the difficulty in part reflects the danger of starting from a scale developed for assessing individual wellbeing in order

[59] Ibid

to derive an assessment of collective wellbeing. Beyond attempts to derive collective profiles from combinations of individual results there have been many attempts to develop scales that capture social data that are assumed to be linked to individual wellbeing, such as income levels, housing provision or adequacy of service provision. In the UK, these have been stimulated to a large extent by the Best Value regime and the work undertaken by the Audit Commission to help local authorities in the systematic gathering of data and collation of this to assess change related to PSA targets. Many consultancy firms and academics have also been working on the development of IT tools to facilitate combining data into single measures. Initial difficulties identified in this area was the lack of accessible statistics at the local level, eg by work done by the CLIP Task Force on Sustainable Development and the PAT 18 Task Force on Local Statistics[60], but since then considerable efforts have been made to make local statistics more accessible.

Susan Hird, in her excellent review of current literature in this area, suggests a model that identifies five different dimensions in the type of social indicators that have been used to define wellbeing: physical wellbeing, material wellbeing, development and activities, social wellbeing and emotional wellbeing.[61]

Another difficulty with the concepts of quality of life, wellbeing and happiness is that their very breadth has led to many different interpretations of what they mean. Hird comments that: 'The difference [between these terms] depends on who and what you read. Some researchers will state that many of the terms are synonymous, others insist that there are fundamental differences.'[62]

The breadth of the concept of quality of life is highlighted in recent work by the Audit Commission, which identified the following key dimensions to the concept in relation to the achievement of sustainable community strategies:

- community cohesion and involvement
- community safety
- culture and leisure
- economic wellbeing
- education and life-long learning
- environment
- health and social wellbeing
- housing
- people and place
- transport and access.[63]

[60] Social Exclusion Unit, *Report of Policy Action Team 18: Better information*, London, Cabinet Office, 1999
[61] Hird S, *Community Wellbeing: A discussion paper for the Scottish Executive and Scottish Neighbourhood Statistics*, Edinburgh, NHS Scotland, 2003
[62] Ibid
[63] Audit Commission, *Local Quality of Life Indicators: Supporting local communities to become sustainable*, London, 2005

Recently there has been considerable criticism that wellbeing has been too closely equated with specific dimensions such as economic growth, or overall level of health, while other dimensions of wellbeing such as family relationships or the strength and safety of communities have been overlooked.

Attempts to redress this imbalance has led to a growing interest in attempting to assess communal dimensions of wellbeing, work stimulated particularly by Putnam's work, which introduced the concept of social capital.[64] It would be difficult to do justice in this paper to the volume of work that has been undertaken over the last decade since Putnam first published his work. However, it is relevant to note that there is still considerable uncertainty over exactly how the concept of social capital can be operationalised in practice and about how in reality it contributes to wellbeing either at an individual or collective level.

Nevertheless, many attempts have been made to translate the concept into tools for measurement. At the level of individuals this involves measures for assessing individuals' subjective experience, such as their awareness of community-level structures, their experience of trust and involvement in their neighbourhood, their sense of optimism and belonging to their area.[65] More objective measures have attempted to map individual networks and the level of involvement with these networks.[66]

At the collective level there have been tools developed for the measurement of community capacity-building. Initial work on mapping local organisations to assess their number, size and membership, and extent of volunteering[67] evolved into a more sophisticated framework for measuring 'community', which identifies 16 different variables grouped into six different areas:

- Individual – factors such as self-determination, concern with public issues and level of volunteering or community attitudes
- Community involvement, horizontal – numbers of community and voluntary organisations, social capital and mutual aid
- Community involvement, vertical – voting turnout, response to consultations, extent and effectiveness of community representation
- Services and economic development – extent and range of contribution to public services, social economy and assets
- Inclusion/diversity/cohesion – including measures of extent to which all sections of the population co-exist harmoniously and co-operate with one another

[64] Putnam R D, *Bowling Alone: The collapse and revival of American community*, New York, Simon & Schuster, 2000
[65] Walker P and Lingayah S, 'Measure for Measure: Indicators for LA21 and indicators for regeneration', *Local Environment News* (University of Westminster), Vol 5 No 3, March 1999
[66] Chanon G, *Measures of Community*, Community Development Foundation, 2004
[67] Chanan G, Gilchrist A and West A, *Involving the Community: Single Regeneration Budget*, London, Community Development Foundation, 1999

- Provision/support/empowerment – community development provision, community and voluntary sector infrastructure and support from partnerships, and public services.[68]

There has also been a growing interest in attempting to operationalise broader notions of happiness or wellbeing at a community level. For example, those undertaken by the CSAE global poverty research group and the ESRC wellbeing in developing countries research group.[69]

The foregoing discussion about the challenge of measuring these dimensions highlights the way in which underlying assumptions at the development stage of outcome measures of this kind help to shape the overall emphasis of the measure, which might or might not reflect the actual preferences of the public or the complexity of the issues involved.

However, there does appear to be a growing recognition that the development of outcome measures of this kind requires public consultation. For example, the NICE guidelines on measuring impact in terms of health and wellbeing for people in mid-life and beyond.[70] This provides a menu of methods for measuring the effectiveness, but also recommends that initial work is undertaken with a range of stakeholders including the public and users of services in order to identify what should be measured. It also recommends the use of a range of methods for capturing information (including surveys, use of routine administrative information, gathering people's experience in more qualitative ways, photographs and videos of evidence) and triangulating this data.

---

[68] Ibid

[69] De Jong J, *Capabilities, Reproductive Health and Wellbeing*, Global Poverty Research Group working paper, 2004

[70] National Institute for Health and Clinical Excellence, *Measuring Impact: Improving the health and wellbeing of people in mid-life and beyond*, NICE, 2005

Table 14: Quality of life, wellbeing and happiness scales and the measurement framework

Quality of life, wellbeing and happiness scales		Stages in the policy cycle				
		Agenda setting	**Task definition**	**Implementation and delivery**	**Outcomes**	**Impact**
Are quality of life, wellbeing and happiness scales:	**Appropriate?** • Are they fit for purpose? • Do they meet relevant methodological standards, including method-specific quality standards?				• Development of appropriate scales is complex and scales developed for assessment of individual wellbeing may not be appropriate for assessing change at a collective or system level	
	Holistic? Do they take into account: • the complexity of the situation? • new public management values such as effectiveness and efficiency? • relevant public value-type values, such as wellbeing, social capital and quality of life?				• Quality-of-life, wellbeing and happiness scales indicate useful ways of assessing wider aspects of 'public value' • These usually involve combining multiple factors, but may give undue weighting to one aspect over another	
	Democratic? Do they allow for: • public involvement? • negotiation between different stakeholders? Are they transparent and accessible? Have they been 'authorised'?				• The technical process of developing scales may make them unsuitable for public involvement. However, different views about what constitutes wellbeing or quality of life usually taken into account in their development	
	Trustworthy? Are the measures and their findings being used appropriately and with integrity?				• An insistence on transparency and a consideration of 'legitimacy and support' will help to make the method trustworthy	
	Generating public value? Does the process of measuring create value in and of itself?				• Use of these broader measures may overcome difficulties of use of more limited measures of outcome	

4. Conclusion

We have tried to set out above a framework that incorporates a set of criteria by which the appropriateness of different measures for assessing public value might be judged. We have then sought to apply this framework to a few examples of different types of measure that might be used at different stages of the policy or programme development cycle.

We accept that this represents only the beginning of a discussion about how the measurement of public value might be undertaken, and have put forward the framework for discussion and debate rather than as a definitive statement.

That said, what we have sought to highlight through the use of the framework is that the process of using measures developed in a somewhat different 'paradigm' to public policymaking is problematic and fraught with dangers. We have also identified the fact that there are discussions already taking place that are very relevant to the whole question of measuring public value, but as yet are not being directly linked to the public value discussions. For example, many of the new models of programme evaluation being discussed and debated in the evaluation field seek to address the issues of stakeholder negotiation, complexity and public engagement that we suggest is required to ensure that measurement systems add to rather than detract from public value. As far as we know, there has not been a great deal of discussion about the implications of the concept of public value in the evaluation community. This is something we hope to address in a workshop in the upcoming UK Evaluation Society conference.

We have also identified some promising leads in addressing some of the issues raised in the area of public opinion polling, contingent valuation, performance measurement, balanced scorecards and in the development of new approaches to the assessment of quality of life, wellbeing and happiness at a collective level. It has been difficult to do justice to any of these areas given the size of the literature available, but the review did indicate some promising leads, which might be followed up in a more extensive exercise than the present one.

Bibliography

Acheson Sir D, *Independent Inquiry into Inequalities in Health Report*, London, HMSO, 1998

Appleby A, Harrison A and Devlin N, *What is the Real Cost of More Patient Choice?*, London, King's Fund, 2003

Appleton J, *Government by Opinion Poll*, Spiked Politics, 2005. Available at www.spiked-online.com/Articles/0000000CA99A.htm

Atkinson T, *Final Report of the Review of the Measurement of Government Output and Productivity for the National Accounts*, London, ONS, 2005

Audit Commission, *Area Profiles Phase II: Quality of life themes*, 2005. Available at http://www.areaprofiles.audit-commission.gov.uk/(mslkl455q3gj5d455k5pr1ua)/StaticPage.aspx?info=25&menu=5

Audit Commission, *Local Quality of Life Indicators: Supporting local communities to be more sustainable*, London, 2005

Benington J, *Creating and Sustaining Public Value in the Health Service: University of Warwick Health Service Partnership Discussion Paper One*, Coventry, Warwick Institute of Governance and Public Management, University of Warwick, 2005

Blaug R, Horner L and Lekhi R, *Public Value, Politics and Public Management: A literature review*, London, The Work Foundation, 2006

British Library, *Measuring Our Value: Results of an independent economic impact study commissioned by the British Library to measure the Library's direct and indirect value to the UK economy*, London, British Library, 2004. Available at: www.bl.uk/pdf/measuring.pdf

CABE, *Physical Capital: How great places boost public value*, London, 2005. Available at http://www.cabe.org.uk/AssetLibrary/2189.pdf

Cabinet Office, *Political Trust and Engagement Seminar Papers: Public assessment and performance*, unpublished, 2004

Chanan G, *Measures of Community*, Community Development Foundation, 2004

Chanan G, Gilchrist A and West A, *Involving the Community: Single Regeneration Budget*, London, Community Development Foundation, 1999

Chen T H and Rossi P H, 'The Multi-Goal, Theory-Driven Approach to Evaluation: A model linking basic and applied social science', *Evaluation Studies Review*, Annual 6, pp38-54, 1981

Cowling M, *Measuring Public Value: The economic theory*, London, The Work Foundation, 2006

De Jong J, *Capabilities, Reproductive Health and Wellbeing*, Global Poverty Research Group working paper, 2004

Denhardt R B and Denhardt J V, 'The New Public Service: Serving rather than steering', *Public Administration Review*, Vol 60, No 6, pp549-559, 2000

Ecosystem Valuation, *Contingent Valuation Method*. Web publication available at www.ecosystemvaluation.org/contingent_valuation.htm

ESRC Global Environmental Change Programme, 'Making Environmental Decisions: Cost-Benefit Analysis, Contingent Valuation and Alternatives', *Proceedings from the Centre for the Study of Environmental Change/Green Alliance Practitioners' Seminar,* held on 23 January 1997 at McKenna & Co, London. Available at: www.sussex.ac.uk/Units/gec/pubs/reps/gralconf.htm

Fishkin J, *The Voice of the People: Public opinion and democracy*, New Haven and London, Yale University Press, 1995

Government Chief Social Researcher's Office, *The Magenta Book: Guidance notes on policy evaluation and analysis*, London, Government Chief Social Researcher's Office, 2003

Guba E and Lincoln Y, *Fourth Generation Evaluation*, Newbury Park, California, Sage Publications, 1989

Hird S, *Community Wellbeing: A discussion paper for the Scottish Executive and Scottish Neighbourhood Statistics*, Edinburgh, NHS Scotland, 2003

HM Treasury, *The Green Book: Appraisal and evaluation in central government*, London, HM Treasury, 2003

Horner L, Blaug R and Lekhi R, *Public Value Final Report: Deliberative democracy and the role of public managers*, London, The Work Foundation, 2006

Kelly G, Muers S and Mulgan G, *Creating Public Value: An analytical framework for public service reform*, London, Cabinet Office, 2002

Lawrence C, *Medicine and the Making of Modern Britain 1700-1920*, New York, Routledge, 1994

Layard R, *Happiness: Lessons from a new science*, London, Allen Lane, 2005

Luskin R, Fishkin J and Jowel R, 'Considered Opinions: The first deliberative poll', *British Journal of Political Science*, 1999

Miller D and Carrier J G, *Virtualism: A new political economy*, New York, Berg Publishers, 1998

Moore M, *Creating Public Value: Strategic management in government*, Cambridge, MA, Harvard University Press, 1995

Moore M H, *The Public Value Scorecard: A rejoinder and an alternative to 'Strategic Performance Measurement and Management in Non-Profit Organizations' by Robert Kaplan*, HCNO Working Paper Series, May 2002

Mulgan G, *Public Value: Physical capital and the potential of value maps*, London, CABE, 2005

National Centre for Social Research, *Deliberative Polls*. Web publication available at www.natcen.ac.uk/natcen/pages/report_deliberative.htm

National Centre for Social Research, *Social Attitudes and Elections*. Web publication available at www.natcen.ac.uk/natcen/pages/or_socialattitudes.htm

National Institute for Health and Clinical Excellence, *Measuring Impact: Improving the health and wellbeing of people in mid-life and beyond*, London, NICE, 2005

Neighbourhood Renewal Unit, *Research Report 7: New Deal Communities, National Evaluation – Annual Report 2003-2004*, London, DCLG, 2003

New B and Neuberger J (eds), *Hidden Assets: Values and decision making in the NHS*, London, King's Fund, 2002

Office of the Deputy Prime Minister, *Public Participation in Local Government: A survey of local authorities*, London, ODPM, 2002

Osborne D and Gaebler T, *Reinventing Government: How the entrepreneurial spirit is transforming the public sector*, Wokingham, Addison-Wesley, 1992

Parsons W, *Public Policy: An introduction to the theory and practice of policy analysis*, Edward Elgar, Cheltenham, 1985

Pawson R, Boaz A, Grayson L, Long A and Barnes C, *Types and Quality of Knowledge in Social Care: Knowledge review three*, London, Social Care Institute for Excellence, 2003

Pawson R and Tilley N, *Realistic Evaluation*, London, Sage Publications Ltd, 1997

Putnam R D, *Bowling Alone: The collapse and revival of American community*, New York, Simon & Schuster, 2000

Ridgeway V F, 'Dysfunctional Consequences of Performance Measurements', *Administrative Science Quarterly*, Vol 1, 1956

Rosser R, 'A Health Index and Output Measure' in Walker S R and Rosser R M (eds), *Quality of Life: Assessment and application*, Lancaster, MTP Press, 1988

Sanderson I, 'Evaluation in Complex Systems', *Evaluation*, Vol 6 No 4, pp433-454, 2001

Scottish Office, *Involving Civil Society in the Work of Parliaments*, Scottish Office. Available at www.scotland.gov.uk/government/devolution/cpsp-00.asp

Smith P, 'On the Unintended Consequences of Publishing Performance Data in the Public Sector', *International Journal of Public Administration*, No 18 Vols 2 and 3, pp277-310, 1996

Social Exclusion Unit, *Report of the Policy Action Team 18: Better information*, London, Cabinet Office, 1999

Spencer L, Ritchie J, Lewis J and Dillon L, *Quality in Qualitative Evaluation: A framework for assessing research evidence*, London, Government Chief Social Researcher's Office, 2003

Walker P and Lingayah S, 'Measure for Measure: Indicators for LA21 and indicators for regeneration', *Local Environment News* (University of Westminster), Vol 5 No 3, March 1999

White C, Lewis J with Elam C, *Citizens' Juries: An appraisal of their role based on the conduct of two women-only juries*, London, Cabinet Office Women's Unit, 1999

Wilmot A, Jones J, Dewar A, Betts P, Harper R and Simmons E, *Public Confidence in Official Statistics: A qualitative study on behalf of the Office for National Statistics and the Statistics Commission*, London, Office for National Statistics, 2005

We provide:
Research
Consultancy
Leadership
Advocacy
Partnership

Registered as a charity no: 290003

First published: December 2006

The Work Foundation
3 Carlton House Terrace
London
SW1Y 5DG

Telephone: 020 7004 7100

Email: enquiries@theworkfoundation.com

Website: www.theworkfoundation.com