

Playing the game of Outcome-Based Performance Management (OBPM). Is gamesmanship inevitable? Evidence from theory and practice

Forthcoming in: *Social Policy and Administration*

Authors

Dr. Toby Lowe and Prof. Rob Wilson, Newcastle University Business School

Abstract

This article presents the case for the need for a re-think in the prevailing orthodoxy of measurement approaches in the governance and management of public services. The paper explores the simplification of complex reality that Outcomes Based Performance Management (OBPM) requires in order to function, and the consequences of such simplification. It examines the evidence for and against the effectiveness of OBPM, and argues that both sets of evidence can be brought into a single explanatory story by understanding the theory of OBPM. The simplification required to measure and attribute 'outcomes' turns the organisation and delivery of social interventions into a game, the rules of which promote gamesmanship, distorting the behaviour of organisations, managers and practitioners who undertake it.

Keywords

performance management; outcomes; outcome based performance management; new public management; information management; payment by results; gaming; gamesmanship

Introduction

The management of the performance of public services is a significant topic for social policy and management scholars. From the 1980s onwards questions began to be raised in the UK about the continuing expansion of the state. This led to a move from what has been described as an era of professional autonomy coupled with bureaucratic systems, to an approach where the focus moved toward efficiency in the production of public services. This became known as “New Public Management” (NPM) (Hood 1991). NPM was later summarized by Ferlie et al. (1996) as 'three Ms': Markets, Managers and Measurement. The measurement element, which forms one of the key planks of NPM, foregrounded a way of thinking in which ‘Accountability’ and ‘Performance Management’ were key and led to significant changes to the way in which public service delivery was organized (Lapsley 2008).

This paper will argue that the current evolution of the measurement element of NPM - Outcomes-Based Performance Management (OBPM), which has become the key mechanism for the implementation of accountability and performance management, turns the performance management of social interventions into a simplified game, which does not deal well with the complex reality of life. (We will generally use the term ‘social interventions’ rather than ‘public services’, as a significant element of this work is devised and delivered by organisations which are not public sector bodies).

The paper will explore the evidence which has been generated in response to the question: is OBPM effective? It will argue that the evidence which demonstrates that targets improve performance data and the evidence which suggests OBPM undermines effective practice are both valid, and that they can be brought into a single explanatory narrative by understanding the theoretical assumptions which underpin OBPM, and the flaws that these contain. It will construct this explanatory framework by exploring the theoretical assumptions which underpin OBPM, and argue that the processes of simplification demanded by the theory of OBPM turn the design and delivery of social interventions into a game which requires certain tactics in order to be played most effectively.

What is Outcomes Based Performance Management (OBPM)?

OBPM is an umbrella term for using “outcomes” as a way of making judgement about the performance and effectiveness of social policy interventions (Lowe 2013). It suggests that the effectiveness of social policy interventions should be judged on the basis of the impact they make in the lives of the people for whom they are designed, and that those people and organisations who deliver these interventions should have their performance rewarded or punished on the basis of whether desired “outcomes” are occurring. (Bovaird 2012, Centre for Social Justice 2011, UK Cabinet Office 2012).

The appeal of OPBM is obvious – it is supposed to focus the attention of those delivering social policy interventions on those whom they serve. As one proponent states: the “clear and compelling answer to the question of ‘Why measure outcomes?’ is: To see if programs really make a difference in the lives of people.” (United Way of America 1996: 4)

During the period in which OBPM initially became popular it represented an evolution in thinking about performance management which moved it on from measuring either the *inputs* into a service (e.g. how many staff are employed to do a task) or the *outputs* of that service - the amount of provision that is offered (e.g. the number of classes offered on a learning programme). Different forms of OBPM have emerged over time. It began in the 1990s as “Outcomes-Based Evaluation”, pioneered, amongst others, by Robert Schalock (see, for example, Schalock 2001), before moving into the broader field of performance management, where it became known as “Management by Results” (Perrin 1998). It is now widely known through Mark Friedman’s programme of “Results Based Accountability” TM (Friedman 2001 - also known as “Outcome Based Accountability”). Most recently, OBPM is the basis of Payment by Results (PbR) and Social Impact Bonds, in which people (or more commonly organisations) are paid by the “outcomes” that they deliver. (UK Cabinet Office 2012, National Audit Office 2015). This is the logical conclusion of the OBPM model, in which the financial rewards for people and organisations are tightly coupled to the delivery of agreed “outcomes”.

Is OBPM effective?

Research into the effectiveness of OBPM approaches has generally fallen into one of two camps. In the first camp, researchers have explored the impact of performance targets on the performance of those who undertake social interventions (mostly public services) (see, for example, Boyne and Chen 2006, Boyne 2010, Kelman and Friedman 2009). This is largely quantitative research, which undertakes large-scale econometric analysis of performance data. It has produced results which have been interpreted as supporting the proposition that OBPM is effective.

In the second camp, researchers have explored the impact of performance targets on the practice of those who commission, manage and deliver social interventions. (See for example: Bevan & Hood 2006b; Perrin 1998; Mayne 2007; van Theeil & Leuww 2002; Rothstein 2008; Newton 2012; Soss, Fording & Schram 2011; Keevers et al 2012). This is largely qualitative research, based on interviews with those who undertake the work. It finds that OBPM distorts and undermines the practice of social interventions – this distortion is often called ‘gaming’ (e.g. Bevan and Hood 2006b).

Let us explore the evidence produced by these two camps in turn.

Evidence supporting OBPM

Our departure point for this exploration will be Boyne's (2010) piece, which asks whether performance management works. (Although, Boyne does not use the term ‘Outcomes Based Performance Management’, it is clear that he is referring to this idea, as he describes “the central purpose of these initiatives has been unchanging: to improve public management and program outcomes” (Boyne 2010: 209) and he describes the three elements of performance management as “selecting indicators, setting targets, and taking action to influence scores on the indicators and the extent of target achievement” (Boyne 2010: 209).

Boyne (2010: 223) concludes that “the balance of the empirical evidence is consistent with the meta-proposition that performance management is associated with better public services.”

In order to reach this conclusion, he draws on a variety of studies, including his own study of the effects of target based performance on Local Authorities (Boyne and Chen 2006). Boyne and Chen (2006) find that:

“The results show that authorities with a target performed better than their peers in the LPSA period and better than themselves in the pre-LPSA period. This evidence supports the view that setting a target on a performance indicator leads to better outcomes, at least as measured by exam results in the education service.” (Boyne and Chen, 2006: 472)

They reach this conclusion by undertaking econometric analysis of the performance of English Local Authorities against different kinds of target. In addition, they analyse the figures to see if they contain evidence of particular types of gaming. They conclude that they do not:

“The insignificant coefficients for this variable imply that target effects have not been achieved by “cream skimming” in the selection of pupils for exams. This does not exclude the possibility of other effects of targets on equity. For example, local authorities may have improved the percentage of pupils achieving five or more GCSEs at grades A*–C by concentrating their efforts on those just below this level, thereby neglecting pupils far above or below the threshold of LPSA success. This is “milking the middle” rather than cream skimming but raises equity issues that require investigation in future studies of target effects.” (Boyne and Chen, 2006: 474)

Kelman and Friedman (2009) come to similar conclusions by undertaking an econometric analysis of performance data relating to NHS waiting time targets. Their starting point is to examine the measurement of results of waiting-time target indicators. These waiting time indicators demonstrate significant improvement during the period in which the Government made waiting times a key performance target for hospitals. Based on previous qualitative evidence about the distorting effects of targets, they then construct and test scenarios within the data which might suggest that such improvements were due to ‘gaming’ rather than genuine service improvement:

“First, we discuss theoretically— using literature from public management, economics, organization theory, and accounting— why one might expect dysfunctional responses to adoption of performance measures in an organization and what the different categories of such distortions might be. We illustrate this with examples of distortions predicted for the English A&E wait-time performance target. Second, we present empirical results, based on econometric analysis of data from all English hospitals during the period 2003–2006, on presence of the predicted dysfunctional effects. We find no evidence of these dysfunctional responses. Indeed, in a number of cases , the sign of statistically significant effects predicted

by those worried about dysfunctional effects went in the “wrong” direction, that is, that better wait-time performance was associated with a lower level of problems predicted by a dysfunctional effects story.” – (Kelman and Friedman 2009: 919)

Furthermore, in addition to challenging the idea that such improvements were due to gaming, the authors identify examples of practice improvements which could plausibly be responsible for the improvement. (Kelman and Friedman 2009: 939)

Drawing together the evidence presented by Boyne (2010), Boyne and Chen (2006) and Kelman and Friedman (2009), we can reach the following conclusions:

Conclusion 1: The combination of increased resources and performance targets for services improves the performance data around those targets

Conclusion 2: Analysis of this performance data does not support the claim that it has been produced via certain types of large-scale gaming

Conclusion 3: The introduction of performance targets may have, in some cases, led to improved practice

We will now consider the range of evidence which highlights the problems associated with OBPM.

Evidence against OBPM

The evidence against OBPM suggests that the adoption of such management techniques changes the practice of those who design and undertake social interventions for the worse. Instead of serving the needs of their clients, managers and frontline workers become focussed on how to produce the required performance information. This evidence is largely based on qualitative interviews with public officials, staff and managers.

A recent example of such research is Soss, Fording & Schram (2011). They conducted a detailed study of the behaviour and perspectives of officials, managers and staff who are delivering the Temporary Assistance for Needy Families (TANF) programme in the United States. This programme is commissioned and performance managed using a PbR framework which disciplines all those involved with the programme, both staff and those in receipt of support.

They found that the OBPM regime:

- Focuses staff time on the production of data, rather than supporting clients (Soss, Fording & Schram 2011: 221)
- Encourages managers to employ people with data-processing skills, rather than those with the skills needed to support vulnerable people (Soss, Fording & Schram 2011: 221)
- Encourage managers to find ways to meet performance targets which are not based on improving service to clients Soss, Fording & Schram 2011: 211/12)
- Encourages staff at all levels to behave perversely, subverting the stated ambition of the programme. (Soss, Fording & Schram 2011: 207/8)

All these effects are summarised as follows:

“In the WT program, performance is the name of the game for local service providers. But organizations typically adapt in perverse ways, and internal contradictions embedded in the NPM work systematically against policy learning and program improvement” (Soss, Fording & Schram 2011: 212)

A similar picture was found in a study of the implementation of Results Based Accountability TM in Australia (Keevers et al. 2012). This study is significant because it is one of the few which has undertaken an in-depth ‘before and after’ look at how the introduction of Results Based Accountability affected the practices of frontline staff within social support organisations. It found that following the introduction of a Results Based Accountability reporting system, staff spend time collecting and analysing data about those young people rather than spending time developing and maintaining the quality of relationships which are the cornerstone of their work with young people (Keevers et al. 2012: 114).

The negative effects of the OBPM regime can clearly be seen from the way in which the workers describe its impact on their practice:

“It’s constantly looking at numbers. Yeah, and the quality and depth of the client contact has really declined in the last couple of months because of the pressure of the new data and

monitoring requirements. We don't get the funding unless we meet the targets. It's really changed the way we work. ... If a young person was having problems with transport or anything like that we would go and meet the client. We would either go to where they were staying and do the assessment there, or we'd take them somewhere where they felt more comfortable, so we might meet them at McDonalds or something like that. Now we can't – we can't do any of that because we have to enter information onto the computer as soon as they come in. And they [funding body] have either booked us an appointment right after or there is not enough space between times to drop them off and pick them up. (reflective discussion)" (Keevers et al 2012: 114)

Such accounts of the way in which OBPM regimes serve to shape the thinking of those who work under them add depth and detail to other research with staff at all levels who have been involved with OBPM mechanisms. Together, this research suggests that in order to produce the required data, people 'game the system' in various ways: creaming & parking, teaching to the test, reclassifying, and falsification of data. These studies find this same 'gaming' activity in a huge variety of policy settings, across a range of different places (Bevan & Hood 2006b – UK, Health Service; Perrin 1998 – USA and Canada, employment programmes; van Theeil & Leuww 2002 – Europe & USA, public services; Rothstein 2008 – USA, education; Newton 2012: UK, employment programmes)

The conclusion that we are able to draw from this range of research is that, when interviewed, managers and staff from a range of policy contexts across the world who are part of OBPM systems tell remarkably similar stories concerning the way in which OBPM impacts on their practice and shapes their thinking. This impact is to the detriment of the quality and effectiveness of the services that they are delivering.

[Bringing the evidence together: creating a single explanatory narrative](#)

How can it be that, across a wide range of policy and geographical contexts, interviews with managers and staff reveal such consistent stories of the dysfunctional impact of OBPM, and yet those who have searched for the evidence of such effects in the performance data itself find no evidence of such behaviour? How can large-scale quantitative analysis suggest that performance data improves for reasons that aren't associated with gaming, whilst interviews with practitioners suggest that gaming is routine?

There seem to be two potential ways to address the conflicting nature of this evidence:

Option 1: To dismiss one or other set of evidence as incorrect

Option 2: To find a single narrative that explains both sets of evidence

Option 1: Dismissal

It is tempting to attempt to dismiss either one set of evidence or the other. Kelman and Friedman (2009: 938) go some way down this road by describing the evidence for gaming within OBPM as “anecdotal”. Having dismissed the qualitative evidence in this way, they construct an explanation for why gaming patterns do not appear in their data: “(1) complementarity across performance dimensions and (2) ways that dysfunctional responses become self-limiting, (3) management behaviors to limit dysfunctional responses.” (Kelman and Friedman 2009: 938)

Similarly, if we wanted to point methodological fingers at the quantitative evidence supporting OBPM, we could start by exploring whether the statistical techniques used have high levels of explanatory power (there are questions about this) and one could also suggest that the attempts to construct scenarios which would suggest gaming are limited in scope and imagination (for example, looking at only alternative scenario per gaming hypothesis to seek evidence of gaming).

However, the weight of evidence on both sides would seem to suggest that such efforts are not a sensible way to proceed. There is enough similarity between the stories told by both types of evidence (and gathered from a variety of places and policy areas) to suggest that each should be taken at face value.

Option 2: Creating a single explanatory narrative

What is required is to create an explanatory narrative that makes sense of both sets of evidence and unites them into a single explanatory framework. We will approach this task by exploring the theoretical assumptions which underpin OBPM, and using this theoretical understanding to reinterpret the evidence about OBPM. These two assumptions are:

Assumption 1: Outcomes can be measured for the purpose of performance management

Assumption 2: Outcomes are created by (and are attributable to) the interventions of particular programmes or organisations

Assumption 1: Outcomes can be measured for the purpose of performance management

We can begin this exploration by asking the question ‘How do we know if an outcome has been achieved?’ Measuring outcomes is a complicated, resource-intensive business. Schalock (2001: 75) identifies the following elements of effective outcome measurement: methodological pluralism, user designed surveys and interviews, the use of control groups and lengthy post-programme longitudinal follow up – 22 months is considered to be a “short” period of follow up (Schalock 2001: 93)

This is because people’s lives, and the context in which they live them, are complex. Understanding the impact of a social policy intervention requires a level of familiarity with the detail of peoples’ lives as they are lived that only comes from intensive research. Understanding an “outcome” means understanding how it fits into the life of the person experiencing it, as each person will have their own perspective on what the desired outcome looks like. This is well illustrated by Widdershiven and Sohl (1999) who use narrative-based evaluation to highlight how the desired outcome of “autonomy for people with learning disabilities” is perceived very differently by each actor within a particular support programme.

However, this is not how “outcomes” are actually measured within OBPM. If organisations or programmes delivering social policy interventions were required to genuinely measure the outcomes of their work on the lives of those they supported, it would cost more to monitor and evaluate those programmes than it would cost to deliver them. As an illustration, Donald Campbell cites research undertaken into the impact of introducing a minimum-income guarantee for American households. Researchers wanted to know whether this achieved positive outcomes for the families, and crucially, whether it impacted on their motivation to look for work. A minimum income guarantee programme was delivered to 300 families, with a further 300 acting as a control group.

The programme itself cost US\$3 million to deliver. The research cost a further US\$5 million (Campbell 1976: 35).

The conclusion to draw from this theory and practice is that genuine research into outcomes is prohibitively expensive. Therefore, rather than undertake rigorous research into whether outcomes have been achieved, organisations/programmes adopt an alternative strategy, which is to use simple, easy-to-collect data to stand as a proxy for genuine outcome information (Friedman 2001: Section 3.7).

We can see the results of this thinking in action in some of the recently developed outcomes-frameworks in the UK. The ASCOT (The Adult Social Care Outcomes Toolkit) uses a simple four question survey format to determine people's needs and desired outcomes. For example, in order to gather information about the desired outcome that "The service user feels he/she has a nutritious, varied and culturally appropriate diet with enough food and drink he/she enjoys at regular and timely intervals" the following question is posed (see Figure 1 Extract from ASCOT questionnaire instrument below by Netten et al. 2011)

Figure 1: Example question from ASCOT toolkit

Similarly, it is why Body Mass Index (BMI) is used as a proxy measure for obesity. Despite the National Obesity Observatory (2009: 3) stating that BMI is a problematic proxy measure because it ignores "factors such as fitness (muscle mass), ethnic origin and puberty" and that it "does not provide any indication of the distribution of body fat and does not fully adjust for the effects of height or body shape" it still recommends using BMI as a proxy. The National Obesity Observatory (2009: 2) states:

"BMI is an attractive measure because it is an easy, cheap and non-invasive means of assessing excess body fat. True measures of body fat are impractical or expensive to use at population level (e.g. bioelectrical impedance analysis or hydro densitometry), and other proxy measures of body fat are difficult to measure accurately and consistently across large populations (e.g. skin fold thickness or waist circumference)."

So, we have quickly moved from a position where large scale rigorous research is required to understand outcomes in people's lives, to one in which OBPM is undertaken using simple data collection via questionnaire or whatever measure can be used to capture information simply and easily. An "outcome" is not a measure of impact in an individual's life. Instead, an "outcome" becomes what is measurable. As Friedman (2001: 3.3) says, "If we had a thousand measures, we could still not fully capture the health and readiness of young children. We use data to approximate these conditions and to stand as proxies for them."

This exploration has shown that the assumption that outcomes can be measured for the purpose of performance management rests on a process of simplification and abstraction. It is not possible to measure outcomes as they are experienced in the complexity of human lives. Instead, people measure what is measurable and call that an 'outcome'.

Let us explore the second assumption:

[Assumption 2: Outcomes are created by \(and are attributable to\) the interventions of particular programmes or organisations](#)

We will begin by asking the question: What is 'an outcome'? An outcome is a snapshot of the state of affairs in the world, as seen from the perspective of a particular person or group of people. The state of affairs in the world – whether that person has a job, whether they have re-offended, or continue to have substance misuse problems – is produced by the interaction of an enormous range of factors. This can be illustrated by looking at obesity. Whether someone is obese, or not, is exactly the kind of complex outcome with which social policy concerns itself. Researchers have mapped the range of factors which contribute to whether people are obese, and how they interact with one another (see Figure 2 Systems Map of Obesity below from Vandebroek et al. 2007: 74)

Figure 2: Systems map of obesity

This map demonstrates the staggering complexity of interactions that lead to 'an outcome'. Not only do each of the individual factors relate in a complex way, but the relationship of each factor to

the whole system is complex. Such systems are described as “non-linear”, meaning “a system whose output is not proportional to its input... Here we can have changes in effects which are disproportionate to changes in the causal elements(s).” (Byrne & Callaghan, 2014: 18). Such systems exhibit “general deterministic chaos where very small variations in the input parameters can generate very different output values in a system of equations.” (Byrne & Callaghan 2014: 19).

Complex systems such as these demonstrate emergent properties. That is, the outcomes that are produced by such systems are not predictable from their starting properties. (Byrne & Callaghan 2014: 21).

We can therefore see that outcomes are emergent properties of complex, non-linear systems. It is untenable to claim that an outcome is created by any one organisation or programme (or even a combination of organisations/programmes). They are the result of the operation of the entire system – a system which not only includes the individual, but also that individual’s interaction with wider society, as Christakis and Fowler (2009) have identified in what they describe as the connected network effect namely: “Your Friends’ Friends Can Make You Fat”.

The key feature of complex systems is that they produce non-repeatable results (Snowden 2003). A person or organisation might act in exactly the same way on two different occasions, but their actions will lead to completely different outcomes, because of the way in which their activities interacted with the whole system. Therefore the same intervention delivered to two different people, or to the same person, but at different times, may well have a completely different outcome. As a consequence of this complexity, outcomes cannot be reliably attributed to interventions (Mowles 2014).

It may be thought that regression analysis can address this issue. Indeed, this is precisely the solution that Schalock (2001: 68) recommends to the problem of attributing outcomes to particular interventions. However, this is not a viable answer, as regression analysis can only use known

elements to which variables were assigned at the start of a study – meaning that all the important variables would need to be known in advance. Secondly, even if all the important variables could be known, using regression analysis within complex systems is exceptionally challenging:

“The development of regression models which have so dominated quantitative social sciences of a non-experimental form... is completely predicated on straightforward linear modelling and efforts to get beyond this by deploying non-linear equation systems...have been, with some exceptions, generally unsuccessful. The blunt point is that non-linearity is a product of emergence. We need to start from emergence and develop a science that fits that crucial aspect of complex reality.” (Byrne and Callaghan 2014: 6/7)

How then does OBPM deal with the problem of attributing outcomes to causes in complex systems?

As with the previous assumption, the answer is by a process of simplification – by using simple (linear) models to map cause and effect. As Bovaird (2012: 6) points out “More generally, the attribution problem tends to be tackled (where it is not simply ignored) through reference to ‘cause-and-effect chain’ models.”

Rather than seeing outcomes as emergent properties of complex systems, OBPM conceptualizes them as products of simple, linear “programme logic” models. Schallock and Bonham (2003: 231) give a classic example of a Programme Logic model as the recommended way of conceptualising the process of creating outcomes. This is shown in Figure 3 below:

Figure 3: A programme logic model template

This simplification is required in order to achieve OBPM’s management objective: to identify and reward those people and organisations that are producing the desired results. A logic model enables outcomes to be viewed as the product of a particular sequence of interventions in the life of a person or group. Outcome X is the product of undertaking intervention Y on person or group Z. This is a linear model. As well as following the line forward from inputs to outcomes, the causal line can (and should) be read the other way, working backwards from desired outcomes to the interventions which create them. As Friedman (2001: 1.1) says ““Results decision-making uses results (the desired conditions of well-being) as the starting point for making decisions. It is a business-like process that

starts with ends and works backwards to means. It first defines success in measurable terms and uses those measures to gauge success or failure.”

There is some recognition within OBPM thinking that the real world does not function in these linear ways. Mayne, in his study (2007: 95), highlights that “Outcomes are by definition results over which organizations do not have complete control”. It is also recognised by Friedman (2001: 1.1): “the more important the performance measure... the less control the program has over it. This is a paradox at the heart of doing performance measurement well”.

However, the response from OBPM is to say that lack of control doesn’t matter. As Friedman (2001: 1.1) says, when discussing what people should be held accountable for “Don't accept lack of control as an excuse... If control were the overriding criteria for performance measures then there would be no performance measures at all.” From this, we can see that the overriding priority is the process of simplification – whatever it takes to simplify reality until it fits the OBPM model, this is what must happen.

Understanding the evidence – a new way to understand ‘gaming’

Unpicking the theoretical assumptions which underpin OBPM enables us to identify that OBPM faces two key problems (1) most outcomes cannot be authentically measured without incurring prohibitive costs and (2) outcomes are not created by, nor attributable to, particular interventions. At the root of both of these problems is a failure to accept the complexity of life as it is lived by real people. OBPM requires simplicity in order to be possible. It demands that outcomes are measurable and attributable, without this, it cannot function, as the UK’s National Audit Office identified in its recent analysis of PbR programmes:

“The nature of PbR means it is most likely to succeed if the operating environment has certain features, for example results that can be measured and attributed to providers’ interventions. If PbR is applied inappropriately there is a risk that either service quality or value for money may be undermined.” (National Audit Office 2015)

OBPM is possible only under a very limited range of conditions, conditions where the outcomes that people experience are easily measurable, and are created within simple, linear systems which enable attribution. Under all other conditions, the world must be drastically simplified in order to make such performance management possible. Rather than limit OBPM to these rare contexts, those who implement OBPM models have, as we have seen, sought to make reality simple. This simplification takes the following form (1) substituting proxy measurements for genuine outcome measures (2) using linear programme logic models for attribution.

We can see this effect clearly documented in Keevers' (2012) description of how organisations undertaking complex practices to pursue social justice had that complexity stripped away under

Results-Based Accountability (RBA) planning processes:

“This fading of social justice practices to the background is perhaps in part due to RBA planning being premised on a representational view of knowledge. Within this ‘representational idiom’, Pickering argues, people appear as shadows of themselves and their practices become abstracted (1995, p. 6). Certainly, in the RBA planning processes we observed, the participants sat at tables using statistical data, graphs and ‘RBA language’ to develop measurable performance indicators for each ‘result’. Such activities shifted the conversations away from detailed, affective accounts of the rough ground of practice, and in this way the participants appeared like Pickering’s ‘disembodied intellects’.” (Keevers et al 2012: 109)

“The specific, situated practice knowledge and attention to young people at risk living well and contributing to community life which expresses the distinctive character of practising social justice at Southern Youth did not appear on any of the ‘results’ lists we witnessed being constructed. The richness, depth and specific character of local practice knowledge was bleached out (Iedema, 2003) during intra-action with RBA representational practices, resulting in final ‘results’ lists that were generalized and indistinguishable from a generic ‘results’ list for any human population.” (Keevers et al 2012: 112)

Understanding this process of simplification enables us to reframe the existing research evidence concerning the effectiveness of OBPM. As we have seen, this empirical evidence largely falls into two camps: firstly, evidence supporting OBPM, which is mainly econometric analysis of large-scale quantitative performance data, and secondly, evidence which identifies the ways in which OBPM processes undermine effective practice, which is mainly derived from interviews with those who undertake this work.

Gamesmanship - playing the Game of OBPM

We can bring together our theoretical understanding of OBPM together with the research evidence to understand why we see the evidence we do. This is a contribution to a growing literature which seeks to understand the wider practice and impact of performance management (see for example, Pollitt 2013 and Lewis 2015). It is an attempt to conceptualise from the breadth of people's lived experience of performance management, escaping the narrow rationalist, technocratic perspective (Lewis 2015: 2).

The process of simplification and abstraction turns the complex reality of life into a simple game (analogous to the way in which a game such as Monopoly is a simple abstraction from real-life capitalism). In order to succeed at the OBPM game people are required to produce the appropriate performance data. Those who produce the appropriate data are rewarded, those who fail to do so are punished.

The desired purpose behind OBPM is to encourage those who deliver under such regimes to produce appropriate performance data by providing effective services to those they support. And, as we have seen from the evidence from Kelman and Friedman (2011) this can happen. However, the way in which the rules of OBPM game are constructed does not favour this way of playing. Instead, the rules favour a different set of tactics. In this section, we will quote extensively from the Soss, Fording and Schram (2011) study, because it has the greatest level of detail on this issue. It is worth remembering, however, that this picture is repeated across many other studies (Perrin 1998, Van Thiel and Leuww 2002, Bevan and Hood 2006 a and b; Mayne 2007, Newton 2012)

The process of simplification pretends that what is measured is an "outcome", and that such "outcomes" are under the control of those who are being held accountable for delivering them. However, those playing the game at the frontline necessarily engage with reality. They know that the genuine impact of the programme is not being captured by proxy measures, and that outcomes are emergent properties of complex systems. Those confronted with the disconnect between

OBPM's simplified rules and the complex reality of life must find tactics to reconcile the two.

Sometimes they struggle with this:

"They say that we're not, how would you say it . . . a social service agency in a sense, like we're a business . . . But at the same time . . . you're working with people who have needs, who have barriers, and bringing the two together is very difficult. [. . .] There's a number game that we have to play. And when you bring that into it, it's hard for me to sit with an individual there; they're telling me that they have all these barriers. For example, they're coming in and they're telling me that they've been evicted from their apartment, they don't have any place to live, they don't have any food, they don't have any clothes. And then here I am as a case manager you have to participate at 40 hours a week. You know, it's just kind of, it's crazy!" (Soss, Fording and Schram 2011: 220)

"While RBA planning tools offer simplification and standardization for some stakeholders, such tools can create confusion and dilemmas for others (Bowker & Star, 1999, p. 293). Practitioners from community organizations struggled to work out how to account for the importance of relationships within RBA planning practices." (Keevers et al 2012: 109)

As a consequence of being held accountable for outcomes which are beyond their control, staff who are involved with the development and delivery of social interventions learn to manage what they can control, which is the production of data. We can see this exemplified in the evidence gathered by Soss, Fording & Schram (2011:209/10):

"Finally, when local actors respond to performance pressures, they also confront "easy versus hard" paths when deciding whether to focus on improving serve to the existing client pool or, alternatively, selecting a client pool that will make it easier to meet performance goals. Evidence from all regions in this study suggests that the latter path is usually seen as easier. Accordingly, creative efforts to innovate are often directed toward reshaping the clientele rather than serving them more effectively"

"In the WT program, serious reforms designed to deal with problems of poverty and work are (not surprisingly) often viewed as difficult to achieve, and their performance effects are usually seen as distant and uncertain. It is far easier to change how one classifies existing activities and counts measured behaviors. As a result, as one local official told us forthrightly, "people game the numbers all the time." In describing efforts to meet the required participation rate, another regional official explained: "You have to do all sorts of things to fill the participation hours. We've got a client who we found out was taking her pastor to church on Sunday. We went out and asked her pastor to sign on saying this was community service. The trick is to find out what people are already doing and find a way to count it as work or community service. This is how you have to do it."

We can model this behaviour to examine the drivers behind the development of this particular set of tactics for playing the OBPM game. This model begins to explore the rational drivers underpinning the development of tactics needed to play the OBPM game well. Using Figure 4 below we can examine the choices faced by staff by placing them along two axes. Along the horizontal axis staff can judge the probability of whether their choices will create an improvement in the results data. The vertical axis concerns the cost of adjustments that they can make. Some changes that they make will be expensive to implement. Others will be cheaper.

Figure 4: Options for improving results data for staff under an OBPM system

The quadrant that staff will rationally seek to occupy will be the quadrant in which the changes they make will have the greatest likelihood of producing the required data, and those which will be cheapest to implement. Hence, Quadrant 3: Low cost/certain impact is the most desirable. These are the choices that will result in the organisation achieving greater financial return and which will keep the overall cost of programme competitive against other organisations who will be tendering for this work.

We can see that all the choices that exist in the 'best' quadrant (3) are those that involve 'gaming' the system. And these are the choices that have been reported by the evidence from OBPM-commissioned programmes (Perrin 1998, Van Thiel and Leuww 2002, Bevan and Hood 2006 a and b; Mayne 2007, Newton 2012). This model demonstrates the way in which people under an OBPM system are driven towards "the alternative logics of Performance Management" as described by Pollitt (2013).

The drivers for making these choices are felt keenly by staff:

“The stress felt by case managers can be traced partly to their belief that performance numbers matter for job security and trajectory. WT case managers make modest wages in a job with few guarantees, and a nontrivial number have previously received welfare themselves. They often struggle to make ends meet and, as a result, tend to view performance through the prism of their own anxieties as breadwinners. Few expect to be “fired” if their numbers drop. But in a system of for-profit contracting, most are keenly aware that performance numbers drive profits, and declining profits could lead their current employer to downsize the staff or even to sell the operation to another company whose retention of old employees is uncertain. At a less absolute level, most expect that if they produce weak numbers, they will be subjected to greater supervision in a way that will make their work more stressful and harder to do. (Soss, Fording and Schram 2011: 221)

This is the nature of playing the OBPM game. The rules of the game create a rationality which favours a set of tactics that have previously been called ‘gaming’. However, all those involved in the system are playing a game. The issue is not about game-playing but about the tactics different players use. A better term for such tactics would seem to be “gamesmanship” – “the use of dubious (although not technically illegal) methods to win or gain a serious advantage in a game” (Wikipedia). [With apologies for the use of gender-specific language – “gamespersonship” is just too ugly and unwieldy to use.]

Each person within the game has to find a way to play it that responds to the underlying rationality of the rules, but which also responds to the complexity of the real world they encounter. The game therefore exists in two separate dimensions: the dimension of simplified rules, and the complex reality of life. (Imagine trying to conduct business affairs using the rules of Monopoly). This means that the game does not ‘make sense’ but still staff must learn to play it well:

“The way we’re able to [stay in business and] help people is by making our measurements on our red and green reports and getting paid, so that we can therefore in return help with childcare and support services [. . .] So the more we make those measurements and those goals, the more we can help candidates. But the more we focus on those [performance goals], the less we’re focusing on the candidates. So, it’s a catch-22.” (Soss, Fording and Schram 2011: 210/11)

It is important to note that “playing the game well” may well involve aspects of delivering the service well. It is not impossible to improve performance data by actually improving the service that is

offered to clients. It is perfectly possible that this occurs, and indeed is likely to be the case in some instances, as the motivations of many of those who do this work are to help those in need:

“case managers are rarely singleminded performance maximizers. More typically, they are ambivalent actors caught in the cross-pressures of competing values, identities, and organizational forces (see also, Watkins-Hayes 2009, 2011). Despite the rhetoric of the “business model,” most express a strong commitment to social service ideals and value their identities as providers who are responsive to clients’ needs. As a result, they express deep reservations about making case decisions based on performance goals.” (Soss, Fording and Schram, 2011: 220)

Therefore the rational driver to develop tactics which maximise the production of good-looking data is tempered by people’s values – their desire to do a good job. However, our model shows that such behaviour – actually delivering the service well – is likely to be expensive, time consuming and have uncertain impact on the data. This is risky tactic because others will be able to construct better-looking data, whilst expending fewer resources.

Using this explanation, we can begin to reconcile what the two different sets of evidence are telling us. One set of evidence tells us that when people are set targets for performance, the data which measures that performance improves. The qualitative evidence gives us a set of reasons as to *how* this performance data improves – it is because those constructing it developed effective tactics to play the game well.

There is one more task remaining in order to construct a story which encompasses both sets of evidence. It must account for why qualitative evidence from staff shows that they use gamesmanship but evidence of such is often absent from the performance data that is recorded. We can explain this by understanding the constructed nature of the quantitative data that has been analysed. The quantitative performance data is large-scale, but flattened through a process of abstraction. It is information from which the complexity has been stripped. It is produced by, and operates within, the logic of the game which has created it. As we have seen, this data construction process happens in many different ways: the decisions of senior managers to hire data manipulators

rather than people with skills in supporting those who are vulnerable (Soss, Fording and Schram 2011: 221), the way in which support is structured to deter or ignore people who are 'difficult' (Newton 2012), decisions about which activities are to be prioritised (Rothstein 2008), and the way that frontline staff classify and record particular pieces of information (Soss, Fording and Schram 2011: 209/10) .

It is perfectly possible for those playing this game to produce performance information which matches the data analysed by Boyne and Chen (2006) and Kelman and Friedman (2009) and yet have undertaken the kind of behaviour uncovered by Soss, Fording and Schram (2011), Perrin (1998), Mayne (2007) and Politt (2013). Both Boyne and Chen (2006) and Kelman and Friedman (2009) use statistical analysis in order to find evidence of the kind of tactics uncovered in the qualitative interviews, but the fact they did not find patterns in the data which match their limited 'gaming' scenarios does not mean that such behaviour was not occurring, as they admit: "We do not claim that dysfunctional effects never occur; we note only that we find no evidence in this case" (Kelman and Friedman 2009; 938).

To reconcile the two sets of evidence, we must say that those playing the game were able to use tactics to create that data in a way which was not apparent to those seeking to interpret a context-stripped version of it. This should not be surprising. Those interpreting quantitative data from which context has been stripped will create new meanings for that data which may be radically different from the meanings given to it by those who created it.

It is also important to understand that our explanatory framework suggests that tactics required to play the game well will look different in different contexts in which OBPM is employed. The greater the level of competition, and the greater the emphasis on PbR, the more the rational drivers of gamesmanship are felt. In these contexts, we would expect there to be less room for people's values to hold sway. In other contexts, with lower levels of competition, with less risk from producing poor-

looking numbers, and with greater ability to contextualise quantitative data with complex stories, the tactics are likely to be different.

Conclusion

We have explored the evidence both that OBPM improves performance data, and that it undermines effective practice. We have been able to reconcile these conflicting areas of evidence by exploring the theoretical underpinnings of OBPM, and, in particular, its reliance on a process of simplification and abstraction in measurement and attribution.

These processes of simplification are required by OBPM in order to try and fit the complexities of life into its models. However, these processes turn the management of social interventions into a game, the rules of which reward the production of data. Playing this game well can involve genuine service improvements, but frequently leads to gamesmanship - tactics which focus on means of data production which do not meet client need.

We have created an alternative way to understand 'gaming'. Gaming is not 'cheating' within a system that otherwise works to improve services for those that need them. The entire OBPM system is a game which is abstracted and simplified from reality. The game measures "outcomes" which are different from how people experience the genuine impact of services and seeks to hold those who play it accountable for things they do not control. As a consequence, they develop tactics which focus their attention on data production, whilst finding opportunities to hold true to their values as best they can.

The theoretical flaws in OBPM mean that this is not a technical problem that can be fixed. It is not a problem that can be fixed by better measurement, or better causal-chain modelling. In order to improve the performance of social interventions, we must move beyond the OBPM approach. If we want people to change their tactics, we must change the nature of the game itself.

Corresponding author

Dr. Toby Lowe – toby.lowe@newcastle.ac.uk

References

- Bevan, G., Hood, C. (2004), Targets, inspections, and transparency. *BMJ* 328: 598.
- Bevan, G., Hood, C. (2006a.), Have targets improved performance in the English NHS? *BMJ* 332: 419–422
- Bevan, G. and Hood, C. (2006b), What’s measured is what matters: targets and gaming in the English public health care system, *Public Administration*, 84, 3: 517–38.
- Bovaird, T. (2014), Attributing outcomes to social policy interventions – “Gold Standard” or “Fool’s Gold”. *Social Policy & Administration*, 48,1: 1–23.
- Boyne, George A. (2010) ‘Performance management: does it work?’ in Walker, Boyne and Brewer (eds) *Public Management and performance: research directions*, Cambridge University Press.
- Boyne, George A. and Chen, Alex A. (2006), Performance Targets and Public Service Improvement, *Journal of Public Administration Research and Theory* 17:3 455–477
- Byrne, D. and Callaghan, G. (2014). *Complexity Theory and the Social Sciences: the state of the art*. London: Routledge.
- Cabinet Office (2012) *Open Public Services White Paper*. London: HM Government
- Campbell, Donald (1976), Assessing the impact of planned social change [pdf]. Available at <https://www.globalhivmeinfo.org/CapacityBuilding/Occasional%20Papers/08%20Assessing%20the%20Impact%20of%20Planned%20Social%20Change.pdf>> Accessed 15th February 2013.
- Centre for Social Justice (2011) *Outcome-Based Government: How to improve spending decisions across government*. London: Centre for Social Justice
- Christakis N, Fowler J. (2009), *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives*, New York: Little Brown and Co.

Ferlie, E., Ashburner, L., Fitzgerald, L., & Pettigrew, A. (1996). *New Public Management in Action*. Oxford: Oxford University Press.

Friedman, M. (2001), *Results Based Accountability Implementation Guide [pdf]*. Available at <http://www.raguide.org/>. [Accessed 28th May 2013]

Keevers, L., et al (2012). "Made to measure: taming practices with Results-Based Accountability." *Organization Studies* 33,1: 97–120.

Kelman, S and Friedman, J (2009). "Performance Improvement and Performance Dysfunction: An Empirical Examination of Distortionary Impacts of the Emergency Room Wait-Time Target in the English National Health Service", *Journal of Public Administration Research and Theory*, 19: 917–946

Lapsley, I. (2008), "The NPM agenda: back to the future". *Financial Accountability & Management*, 24, 1: 77-96.

Lewis, K. (2015) "The politics and consequences of performance measurement", *Policy and Society* 34: 1–12

Lowe, T. (2013), The paradox of outcomes: the more we measure the less we understand, *Public Money and Management*, 33: 3, 213-216

Mayne, J. (2007), "Challenges and lessons in implementing Results-Based Management". *Evaluation*, Vol 13, 1: 87 – 109.

Mowles, C. (2014), Complex, but not quite complex enough: The turn to the complexity sciences in evaluation scholarship. *Evaluation*, 20,2: 160-175

National Audit Office, 2015. *Outcome-based payment schemes: government's use of payment by results*. Available at <http://www.nao.org.uk/wp-content/uploads/2015/06/Outcome-based-payment-schemes-governments-use-of-payment-by-results.pdf> [Accessed 13th July, 2015]

National Obesity Observatory & NHS, 2009. *Body Mass Index as a Measure of Obesity* [pdf]. Available at http://www.noo.org.uk/uploads/doc789_40_noo_BMI.pdf. Accessed on 12th November 2013.

Netten, A. et al (2011), *Adult social care outcomes toolkit v2.1: Main guidance, PSSRU Discussion Paper 2716/3, Personal Social Services Research Unit*. Canterbury: University of Kent.

Newton, B. (2012), *Work programme evaluation findings from the first phase of qualitative research on programme delivery*. Great Britain: Department for Work and Pensions

Perrin, Burt (1998), "Effective use and misuse of Performance Measurement". *American Journal of Evaluation*. 19, 3: 367-379

Pollitt, C. (2013), "The logics of performance management", *Evaluation* 19, 4: 346–363.

Rothstein, R. (2008). *Holding accountability to account: how scholarship and experience in other fields inform exploration of performance incentives in education*, National Center on Performance Incentives [pdf]. Available at https://my.vanderbilt.edu/performanceincentives/files/2012/10/200804_Rothstein_HoldingAccount.pdf. Accessed on 4th February 2013.

Schallock, Robert L. (2001), *Outcome Based Evaluation - Second edition*. New York: Kluwer Academic

Schallock, Robert L & Bonham, Gordon S. (2003), "Measuring outcomes and managing for results", *Evaluation and Program Planning*, 26, 4: 229–235.

Soss, J; Fording R; Schram, S F. (2011) "The Organization of Discipline: From Performance Management to Perversity and Punishment", *Journal of Public Administration Research and Theory*, 21: 203–232

Snowden, DJ. (2003), "Managing for serendipity or why we should lay off 'best practice'" in KM, *Inside Knowledge, ARK Knowledge Management*, 6, 8: 1-8

Thiel, S van. and Leeuw, F. L., (2002) The performance paradox in the public sector, *Public Performance and Management Review* 25,3: 267-281

United Way of America, (1996), *Measuring program outcomes: a practical approach*. United Way of America

Vandenbroeck, P, Goossens, J. Clemens, M. (2007), *Foresight Tackling Obesities: Future Choices – Building the Obesity System Map*. Government Office for Science.

Widdershoven, Guy and Sohl, Carlo, (1999), Interpretation, Action, and Communication: Four Stories about a Supported Employment Program, in Abma, Tineke (ed.), *Telling Tales: On Evaluation and Narrative, Advances in Program Evaluation, Vol. 6*, JAI Press, pp. 109-128

Wimbush, E. (2011), Implementing an outcomes approach to public management and accountability in the UK—are we learning the lessons? *Public Money & Management*, 31, 3: 211-218

Figure 1

3. Thinking about the food and drink you get, which of the following statements best describes your situation?

Please tick (☑) one box

I get all the food and drink I like when I want

I get adequate food and drink at OK times

I don't always get adequate or timely food and drink

I don't always get adequate or timely food and drink, and I think there is a risk to my health

Figure 2

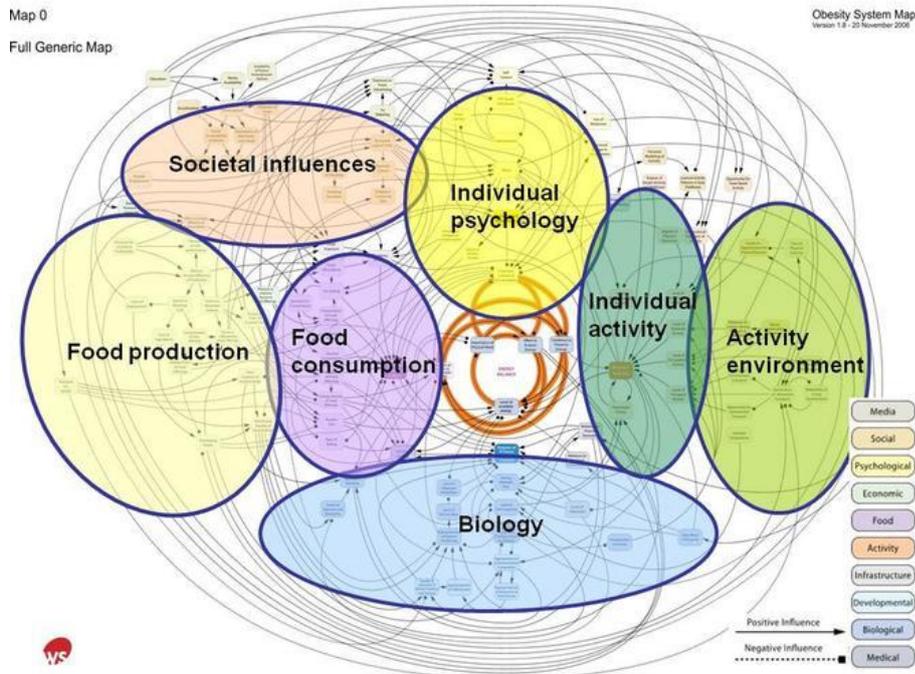


Figure 3

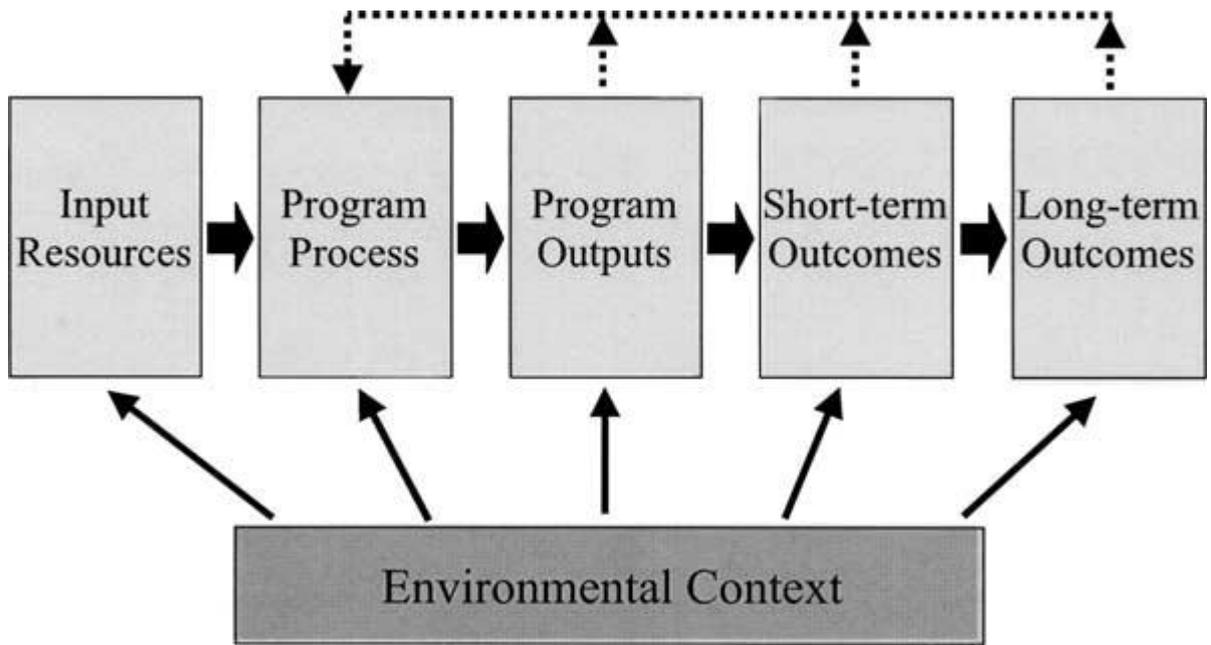


Figure 4

