

Measuring Outcomes

A Review of Academic Literature

Boll, Karen; Oats, Lynne

Document Version
Final published version

Publication date:
2014

License
CC BY-NC-ND

Citation for published version (APA):
Boll, K., & Oats, L. (2014). *Measuring Outcomes: A Review of Academic Literature*. Tax Administration Research Centre, University of Exeter Business School, United Kingdom. TARC Discussion Paper No. 007/14 [http://tarc.exeter.ac.uk/media/universityofexeter/businessschool/documents/centres/tarc/publications/discussion_papers/TARC_Measuring_Outcomes_Discussion_Paper_\(1\).pdf](http://tarc.exeter.ac.uk/media/universityofexeter/businessschool/documents/centres/tarc/publications/discussion_papers/TARC_Measuring_Outcomes_Discussion_Paper_(1).pdf)

[Link to publication in CBS Research Portal](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us (research.lib@cbs.dk) providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 14. Aug. 2022





Tax Administration
Research Centre

Measuring Outcomes
TARC Discussion Paper
007/14

Karen Boll
Copenhagen
Business School

&

Lynne Oats
University of Exeter
Business School

Assisted by:
Carlene Wynter



HM Revenue
& Customs



HM TREASURY

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

The Tax Administration Research Centre undertakes research on tax administration in order to strengthen the theoretical and empirical understanding of tax operations and policies. The Centre is operated in partnership by the University of Exeter and the Institute for Fiscal Studies. It is funded by the Economic and Social Research Council, Her Majesty's Treasury and Her Majesty's Revenue and Customs. The research of the centre is multidisciplinary and the research team includes economists, accountants, experimentalists and psychologists.

For more information about the activities of the centre see:
<http://tarc.exeter.ac.uk>

**Measuring Outcomes: TARC Discussion Paper:
A review of academic literature**

Introduction

This review has been written to coincide with the 2014 publication by the OECD Forum on Tax Administration report *Measures of Tax Compliance Outcomes*. The aim is to canvass a selection of academic literature that is helpful in understanding the challenges of outcome measurement in other spheres, given that limited academic research has taken tax administration as a specific focus.

The rise of performance measurement (PM) systems in the public sector has been widely acknowledged as a significant international trend; a key feature of the New Public Management movement. This trend has not gone unnoticed within tax administrations. Increasingly, PM systems are being used to account for ‘production’ within tax administrations.

Although PM systems are common in tax administration, the academic literature on this is scarce. This review opens by highlighting central points from the few academic studies of PM in tax administration. Then the review briefly describes studies which show challenges in using specifically *output-based* PMs—which is the most common form of PM used in the Public Sector. This then leads to a review of the academic literature on the emergent trend of using *outcome-based* PM in the Public Sector, the subject of the OECD study to which this review seeks to contribute.

A key objective of this review is to show that while it is useful to supplement *output-based* PMs with *outcome-based* PMs, research also shows that *outcome-based* PMs have their own drawbacks—or what might be called ‘unintended consequences’. These are often hidden when the focus is on logic planning, steering and controlling with PMs. In this approach, focus is rightly on best practice, international standards and/or guidelines. What the academic literature can add to this is that the use of *output-* and *outcome-based* PMs is far from linear and straightforward, but rather can lead to unexpected challenges and outcomes.

***Output-based* Performance Measurement in Tax Administration**

The studies on *output-based* PMs in tax administration have—curiously—been conducted in Asia-Pacific and African tax administrations.

Grandcolas studied management strategies of Asia-Pacific tax administrations. He argues for the use of PMs which relate to the *output* to be produced. Such PMs

relate to the “number of tax inquiries handled per person years” or the “time taken to answers inquiries by e-mail” (Grandcolas, 2006: 10). According to Grandcolas, such a focus enables clear measures of performance and progress. Grandcolas’ approach, by focusing on concrete measurable, is appealing, yet, much research and practical experience has shown that this approach to performance measures is prone to gaming and manipulation (see later in this review), and it is too simplistic: Accomplishing internal tasks does not necessarily lead to the desired external outcomes.

Cantens et al (2010) studied the implementation of performance contracts in the Cameroon Customs authority. There, contracts made between the General Director and front line officers, were predominantly implemented because of accusations of corruption and their aim was to set up clear task objectives (outputs) for customs officers. The study evaluates the use of these individual performance contracts as positive; the outcome has been lower corruption, higher revenue collection, and shorter clearance time (ibid: 70). Although evaluating the system positively, Cantens et al. also have their reservations. They observe that the system of performance contracts had only been in place for a few months and also they comment on the fragility of the system, as they acknowledge that it is in fact “..impossible to establish a reliable and constant measurement system” (Cantens et al (2010): 69).

A third example of research into PM in tax administration is the work by Muñoz (2013) who studied the implementation of management by objectives in the Cameroonian tax administration. In this PM system, tax inspectors’ performance is measured against certain aims for revenue collection. In this case, it was the *potential for revenue* that served as a yardstick of performance for the inspectors and this potential for revenue was broken down into revenue quotas for each tax unit and their respective staff. Hence, each tax unit and each member of staff were assigned targets for how much revenue they ought to collect (Muñoz, 2013: 43). While rationally contrived, Muñoz’s study shows that the goal of *potential for revenue* had unintended effects as this ruled out other legitimate goals such as getting more tax payers registered—which was an activity that did not generate revenue immediately.

The findings of these articles are interesting because they relate specifically to tax and customs administrations and because also they show examples of the most widespread and commonly used *output-based* PM, albeit in developing countries. For instance, number of tax inquiries handled, processing or clearance time and quantitative revenue targets. All of these *output-based* PMs are also widespread in tax administrations in the global North.

In order to understand more clearly what the problems and challenges are with these *output-based* PMs, we need to move to the vast literature criticising output-based measures in more general terms, drawing on experiences in other ares of government. This is important, because while tax authorities have unique roles and responsibilities, they also share many common characteristics with other

governmental agencies and there are valuable lessons to be learned from examining experiences in those other agencies.

Deficiencies of *output-based* Performance Measures

The academic literature that criticizes output-based PMs is vast and this review will therefore focus only on a number of seminal studies.

Hood and Bevan's research provides a key contribution to this research area (Bevan and Hood 2006; Hood 2006). These authors show that since the mid-2000s it becomes customary to establish public service (output) targets. For example, this is done within the public health sector where targets are set for minimizing the number of patients waiting more than 12 months or more for surgical operations, or targets may be set for the reduction of the number of people living on the streets (Hood, 2006: 516). The UK government can provide evidence that these targets have been met, yet, Hood questions to what extent the reported performance improvements reflect actual improvement (Hood, 2006). Hood and Bevan's point is that often strategic behavior surrounds such target systems whereby the civil servants perhaps meet the set requirements, but does it via a diminution or even neglect of quality, (Bevan and Hood, 2006: 521). The results are that civil servants 'hit the target', but they 'miss the point' (ibid. 521).

Hood and Bevan also describe the "threshold effects" which refer to the effects of targets on the distribution of performance among a range of, and within, production units (Bird et al. 2005), putting pressure on those performing below the target level to do better, but also providing a perverse incentive for those doing better than the target to allow their performance to deteriorate to the standard and more generally to crowd performance towards the target. Such effects can unintentionally penalize agents with exceptionally good performance but with a few failures, while rewarding those with mediocre performance crowded near the target range (p521).

A similar questioning of *output-based* PMs is put forward in Eterno and Silvermans' analysis of the use of activity based targets in New York City Police Department (2012). They show that police-men's daily work is directed by strict PMs, and that the police worry so much about meeting these, that lowering the crime (the actual outcome) has become a secondary issue.

Meyer (1994, p. 103) concludes: "The long held view of what gets measured gets done has spurred managers to react to intensifying competition by piling more and more measures on their operations in a bid to encourage employees to work harder. As a result, team members end up spending too much time collecting data and monitoring their activities and not enough time managing."

Newberry & Pallot (2004) examine elements of New Public Management, although not specifically performance measurement, in New Zealand central

government departments and find accounting based financial management incentives, instead of improving efficiency and effectiveness, had the opposite effect, eroding resources and constraining managerial freedom.

De Bont & Grit (2012: 497) write: “Performance indicators give a poor representation of the complexity, variety and dynamics of public service production and can create effects that are more perverse than beneficial. Because it is hard to define performance of professional work, performance measurement systems tend to grow into complex systems”

Van Thiel and Leeuw (2002) discuss unintended consequences of performance measurement identifying a performance paradox ie the weak correlation between performance indicators and performance itself. The paradox is about the reports of performance rather than the performance itself.

Finally, De Bruijn (2007) identifies a number of perverse effects arising from public sector performance measurement. According to him, it provides an incentive for strategic behaviour, as described by Osborne and Gaebler – ‘gaming’. It potentially blocks innovation, particularly when the performance measurement system is linked to financial rewards for example bonuses. It may inhibit ambition. Performance measurement can actually obscure real performance, particularly when information is aggregated so that causal connections between effort and performance become lost. A divergence of meanings may arise between internal actors and external actors, and the greater the distance between them, the more likely the external actor is to believe the measures to be unambiguous. Performance measurement may actually drive out a professional attitude because the organisation tends to concentrate on the well-defined tasks.

De Bruijn (2007) further develops five laws of performance measurement.

1. the law of decreasing effectiveness, which states that where a performance measure has high impact, for example if it is used for ranking purposes, it will provide strong incentives for perverse behaviour.
2. the law of mushrooming, which states that performance measurement systems have a tendency to become bloated, through refinement of existing measures and indicators, creation of new rules and explanations for measuring and rules for dealing with exceptions.
3. the law of collective blindness, whereby managers and new staff, who both have less knowledge of grassroots processes, are likely to suffer from myopia; belief that reported performance and actual performance are more highly correlated than they in fact are. This results in a “peaceful equilibrium between managers and professionals. Management is satisfied with the impression that its objectives are being achieved. The professionals are satisfied with not having to put up with managerial interference, or even with being rewarded.” (p45)
4. the law of preservation of perverted systems, noting that once a performance measurement system has become embedded in an organisation, it is not a simple matter to abolish it or even phase it out. He

posits that this is because systems have a ritualising tendency in organisations, and also that external owners of the measures have an interest in maintaining them.

5. the law of decreasing political attention: once performance measurement is institutionalised, political attention declines and it is more difficult to make a case for abolition of particular measures.

To recap, what all the studies demonstrate is that an impact of *output-based* PIs can be that civil servants tend to care more about meeting their performance goals than about doing their job in a sound manner. Hence, stand-alone *output-based* performance measures can have strong gaming and ‘perverse’ side effects.

Outcome-based Performance Measures as a rescue?

One way to avoid focussing solely on *outputs*, i.e. the concrete activities, and to avoid the unwanted manipulation that often follow this, is to use *outcome-based* or *result-based* PM systems. In these systems focus is on the aims for outcomes and results which the given institution wishes to generate. Importantly, these results are “any agency-produced effect on individuals who do not work for that agency” (Swiss, 2005). Such an effect could for instance be that the general customer satisfaction increases or that the tax gap lowers. The advantage is that the public sector institutions working within this governance frame will be more outcome/result oriented. They will be pushed to become more reflective in choosing the right (internal) activities to achieve the right (external) results.

Increasingly, the concept of *nudging* is also connected to this governance frame as the public authorities gently nudges, pushes or directs the decisions of the citizen in ways that they see appropriate (Thaler and Sunstein, 2008). The point is that performing any activity (e.g. inspection, control, service, guidance or information), only carries value if it adds to the fulfilment of the outcomes.

In general, all government administrators across the public sector are under pressure to demonstrate that the *outcomes* of their programmes are effective and meet the expectations of stakeholders (Dubnick & Fredrickson (2011). As Crandall (2010) observes, without a comprehensive performance measurement system, achievement of goals will be difficult and transparency lacking. Hence, steering towards outcomes is a necessary feature of modern public administration.

However, as Dubnick & Fredrickson (2011: 38) note: “[m]easures of performance, if based upon reliable data and kept up to date, may tell the organization in a general way how it is doing. As guides for incremental program adaption, reliable performance data are helpful. But such measures seldom tell an organization what to do” (ibid: p32). Dubnick & Fredrickson (2011: 38) further state that “despite the envisioned benefits of [performance measurement systems], their implementation trajectories are far from linear and straightforward: various difficulties may arise along the way that lead to different outcomes or even to abandonment of the system.

Hence, although outcome-based PMs are advocated as recommend best-practice, and there is a clear trend towards their increased use, the academic literature shows these often have drawbacks or unintended consequences. The studies below highlight the unexpected challenges and consequences which may arise from using outcome-based PMs.

Challenges and unintended consequences of *outcome-based* PMs

In designing outcome based PM systems, it is important to be cognizant of the potential pitfalls. Some of these are outlined below by reference to various academic studies.

- High level of controversies within the organization

Implementation of any new system, particularly PM systems, can lead to controversies within an organisation. In particular, problems can arise where managers used to dealing with output measures are uncomfortable with outcome measures. Arnaboldi & Azzone (2010) investigate the implementation of a performance measurement system in Italian universities. The authors provide a discussion of the difficulties of implementing PM systems in public sector organisations more broadly. From this study, their point is that using PM leads to controversy within the organizations. In these controversies, diverse actors enact different tactics to trying to reach their goals. Instead of seeing this as a (negative) problem, the controversies ought to be a priority for management because these can positively revitalize the interest in and engagement with the PM systems.

- The attribution problem

Bovaird (2014) describes the attribution problem connected to outcome-based PMs, referring to the 'cause-and effect' problem in determining the causes of outcomes, as opposed to the factors which are merely associated with the outcomes. It is simply difficult to determine *what* influences a specific outcome.

This is also highlighted by Boyne and Law (2005) who, in their study of the first generation of UK Local Public Service Agreements, state that “[m]easuring outcomes is complex, and is more so in some service areas than others. A difficulty for some public service providers is that outcomes are strongly influenced by external factors that they are not able to control” (260). They also advocate clear articulation of objectives so that suitable and robust PMs can be identified and where appropriate a basket of indicators rather than one should be used, in order to capture the whole picture.

Mant (2001) also comments that a central weakness of using outcome-based PMs is that other factors influence them. Looking at health care, he argues that it is not only the quality of the care that determines good health as an outcome. Nutrition, environment, lifestyle and poverty also influence health outcomes.

These studies all point to the ‘attribution problem’ which is an inevitably problematic feature of outcome-based PMs.

Law (2013) in her recent study of Welsh outcome agreements, observes that outcomes are often achieved with others in co-production, for example meeting recycling targets may depend on the “willingness of residents to sort, save and put out their recycled materials appropriately”. In a tax context, coproduction of outcomes occurs for example through the actions of intermediaries; accountants, lawyers and others acting on behalf of taxpayers. She also notes that in some cases external factors ‘may be so influential that the outcome is felt to be outside of the control of the organisation, so service managers may as well sit back and see if their numbers come up’.

- **Deprofessionalization**

Adcroft and Willis (2005) warn that the most likely outcome of using (outcome-based) PMs in public administration is deprofessionalisation of public service workers. Their argument is built on the assumption that the use of PM systems will ‘commodify’ relationships within the public sector and that this will affect the ethical distinctiveness of the civil servant and his or her professionalism.

- **Processes are more important than outcomes**

Van Ryzin (2011) argues that the focus on outcomes may be problematic because trust in people and trust in institutions of authority often depends more on process (such as fairness and equity) than on outcomes.

- **Distortion to organizational control**

Smith (1993) focusses on outcome-based PMs. First he shows that there is a widespread dissemination of outcome performance data in the public sector. Second, he outlines the potential distortions induced by an excessive reliance on outcome-based PMs. By analysing concrete examples of outcome-based PM, he shows the significant dysfunctional consequences on control by using performance data.

- **Dysfunctional consequences of publication**

Mason and Street (2011) highlight the need for caution in publishing outcome data in their study of the health sector in the UK and the US. They suggest that media coverage of ‘poor’ performance can lead to a decline in public trust, which in the case of tax administrations could conceivably lead to a decline in one of the key outcomes: securing maximum voluntary compliance. The authors offer five pointers:

1. recognise that outcome statistics may be constructed from data collected for other purposes and need to be treated with caution;
2. recognise that different users have different information needs;
3. consult with target groups in the development of valid quality indicators;
4. understand users’ modes of access to information (although this may be less important in a government agency where users do not have choice of providers);
5. resist the temptation to over-simplify.

Smith (1995) developed a typology of the main types of dysfunctional consequences arising from publishing outcome data in the content of the health service.

To recap, many studies have focussed on the potential (negative) consequences of outcome-based PM systems in the public sector. Together they point at different caveats. These are all important to keep in mind when implementing and developing outcome-based PM systems with in tax and customs administrations.

Conclusion

In this TARC Discussion paper “Measuring Outcomes” we have reviewed a number of the central academic studies dealing with output- and outcome-based PMs.

Three findings are central to tax administrations:

- Output-based PMs are the most widespread indicators in the Public Sector today. However, the academic literature has convincingly argued that these indicators can have strong gaming and ‘perverse’ side effects. This is also shown in studies of tax administrations.
- Outcome-based PMs are newer—but also becoming a more widespread form of indicator in the Public Sector. While these remedy some of the challenges of output-based PMs, outcome-based PMs carry their own ‘unintended consequences’ and challenges.
- Tax authorities have unique roles and responsibilities in our societies. Yet, they also share many common characteristics with other governmental agencies and there are valuable lessons to be learned from examining experiences in those other agencies.

References

- Adcroft, A and Willis, A. (2005) "The (un) intended outcome of public sector performance measurement." *International Journal of Public Sector Management* 18 (5) 386-400.
- Arnaboldi, M. and Azzone, G. (2010) "Constructing performance measurement in the public sector." *Critical Perspectives on Accounting*, v21 pp 266-282
- Barman, E. & MacIndoe, H. (2012) Institutional pressures and organisational capacity: the case of outcome measurement. *Sociological Forum* 27(1) pp 70 – 92.
- Boyne, G. and Law, J. (2005) "Setting public service outcome targets: lessons from local public service agreements." *Public Money and Management* 25 (4) 253-260
- Bevan, G. and Hood, C. (2006) "What's measured is what matters: targets and gaming in the English Public Health Care System." *Public Administration* 84 (3), 517-538.
- Boyne, G. and Law, J. (2005) "Setting public service outcome targets: lessons from local public service agreements." *Public Money and Management* 25 (4) 253-260
- Boviard, T. (2014) Attributing Outcomes to Social Policy Interventions – 'Gold Standard' or 'Fool's Gold' in Public Policy and Management, *Social Policy and Administration* 48(1) pp 1-23.
- Cantens, T., Raballand, G., & Bilangna, S. (2010). Reforming Customs by measuring performance: a Cameroon case study. *World Customs Journal*, 4(2), 55-74.
- Cavulluzzo, K.S. and Ittner, C.D. (2004) Implementing performance measurement innovations: evidence from government, *Accounting Organizations and Society*, 29, pp 243-267.
- Coglianesi, C. (2012) Measuring Regulatory Performance: Evaluating the Impact of Regulation and Regulatory Policy, OECD Expert Paper No 1.
- Collier, P.M. (2006) In search of purpose and priorities: Police performance indicators in England and Wales, *Public Money and Management* 26(3) pp 165 – 172.
- Crandall, W. (2010) "Revenue administration: performance measurement in Tax Administration" Fiscal Affairs Department, IMF Technical Notes and Manuals

- De Bont, A. and Grit, K. (2012) "Unexpected advantages of less accurate performance measurement. How simple prescription data works in complex setting regarding the use of medications." *Public Administration* 90 (2) 497-510
- De Bruijn, H. (2007) "*Managing Performance in the Public Sector*" Routledge, Oxford.
- Dubnick, M., and Fredrickson, H. (2011) "*Public Accountability: performance measurement, the extended state, and the search for trust.*" National Academy of Public Administration, Kettering Foundation.
- Eterno, J., & Silverman, E. B. (2012). The crime numbers game : management by manipulation. Boca Raton: CRC Press, Taylor & Francis Group.
- Fisher, C. and Downes, B. (2008) Performance measurement and metric manipulation in the public sector *Business Ethics: A European Review* 17 (3), 245-258
- Grandcolas, C. (2006). Management of the VAT - Improving the Level of Compliance Using Performance Indicators. *Asia-Pacific Tax Bulletin*, 12(1), 6-18.
- Ittner and Larker (1998) Innovations in performance measurement: trends and research implications *Journal of Management Accounting Research* 205-238.
- Law, J. (2013) Do Outcomes Based Approaches to Service Delivery Work?, Centre for Advanced Study in Public Policy, University of South Wales. Available at: <http://caspp.southwales.ac.uk/media/files/documents/2013-06-24/publication-local-authority-outcome.pdf>
- MacBryde, J.; Paton, S.; Bayliss, M.; and Grant, N. (2014) Transformation in the defence sector: The critical role of performance measurement. *Management Accounting Research*, 25, pp 157 - 172.
- Mant, J. (2001) Process versus outcome indicators in the assessment of quality of health care, *International Journal for Quality in Health Care*, 13(6) pp 475-480.
- Meyer, C, (1994) How the right measures help teams excel, *Harvard Business Review*, 72(3) pp 99-122.
- Muñoz, J.-M. (2013). The revenue imperative in Camerron. Administrative dilemmas. In T. Cantens, R. Ireland & G. Raballand (Eds.), *Reform by Numbers* (pp. 37-57). Washington DC The World Bank.
- Smith, P (1993) Outcome-related performance indicators and organizational control in Public Sector *British Journal of Management* 4, 135-151

Smith, P (1995) On the Unintended Consequences of Publishing Performance Data in the Public Sector, *International Journal of Public Administration*, 18, pp 277-310.

Swiss (2005) A Framework for assessing incentives in Results-Based Management, *Public Administration Review*, 65(5).

Van Ryzin, G. (2011) Outcomes, Process, and Trust of Civil Servants, *Journal of Public Administration Research and Theory*, 21, pp745-760.

Van Thiel, S. & Leew, F. L. (2002) "The Performance Paradox in the Public Sector" *Public Performance and Management Review* 25(3) pp 267 – 281