**Holding on tight – NPM and the New Zealand Performance Improvement Framework**

**Abstract**

The New Zealand Government’s Performance Improvement Framework (PIF) is aimed at public sector organizational improvement. The PIF exemplifies how the characteristics of NPM challenge the evolution of performance systems attempting to go beyond performance management. The paper provides an empirical example of the operation of the PIF, using a theoretical frame derived from the literature. The analysis suggests that the PIF design does not yet mitigate the effects of gaming, synecdoche and reputation-protecting behaviour. The common graded ratings continue to hold the attention of those to whom the PIF is important, acting as a proxy for the whole performance of the organisation. This is a case study of how, despite promising adaptations, performance improvement efforts remain hampered by NPM limitations.

Key words: public organization performance improvement, performance management, NPM, New Zealand

Summary at a glance:

New Zealand Government’s Performance Improvement Framework (PIF) is aimed at public sector organizational improvement. This is a case study of how, despite adaptations, performance improvement efforts remain hampered by NPM limitations.

**Introduction**

Public management reform continues unabated in most countries where NPM-type approaches have resulted in target-based performance management (Boukaert and Halligan 2008, Boyne 2010, Moynihan 2008, Pollitt and Bouckaert 2017, 2011, Van Dooren and Van de Walle 2011, Van Dooren et al. 2015). Academic treatment of performance management and improvement is now vast, with numerous compendiums devoted to the topic (Bouckaert and Halligan 2008, Mizrahi 2017, Moynihan 2008, Pollitt 2017, Talbot 2010, van Dooren and van de Walle 2008).

Globally performance management approaches have been devised to meet the needs of central government departments and to begin to capture performance at a systems level. Experiments in Canada, UK, Norway, Australia and South Africa, for example have all utilized forms of ‘capability reviews’ involving attempts to introduce a current performance rating or view alongside an element of looking towards a ‘future’ desired state (Allen et al. 2017). The demands of political accountability at the system level have resulted in a variety of undertakings to link bureaucratic activity with public outcome-oriented results. Audits, management reports, administrative data reporting and project reviews all form a part of a cluttered system of performance management approaches that have not sufficiently met the need for managing strategic issues that have a present and future impact. They do indicate however, increasing system complexity and an underlying requirement for performance governance that balances both organizational operational level requirements with inter-organizational and multi-stakeholder issues at the system level (Halligan et al. 2012).

This study examines a theoretically informed, practice-based case of New Zealand where public sector reform within an NPM context has been underway for thirty years, and where recent shifts show a public management system which is not entirely NPM in its characteristics and might be seen as an emerging governance hybrid (Osborne 2010, 2013). The Performance Improvement Framework (PIF) was progressively added to an NPM-inspired regime of performance management from 2009 with a specific mandate to help ‘senior leaders in the State Services lead performance improvement in their agencies and across the system on a four-year time horizon’ (SSC n.d).

The key questions the article explores are: what elements of NPM-style performance management impact on efforts to shift to performance improvement to serve the needs of a complex governance regime? To what extent does the New Zealand PIF mitigate against the embedded effects of the NPM approach? We argue that the PIF demonstrates problems emanating from the continuing influence of NPM and that the adaptations of the PIF have not resolved the tensions and paradoxes of performance management (de Bruijn 2009) despite their new dynamics. Evolution of the PIF *has* introduced dialogue and opportunities for learning for the future while the NPM-style measurements anchor the process in a backwards-looking view of performance against static expectations.

This paper contributes to the literature in three ways. As a well-known exemplar of early and continued use of NPM, New Zealand’s creation of and continued evolution of a performance management system is an important example to study. While other comparable system performance management efforts have desisted, New Zealand’s PIF has demonstrated longevity and stability. Second, despite having been used since 2009, thisis the first significant analysis of the PIF based on empirical evidence outside of grey literature. Third, the tensions at play within the PIF demonstrate the issues associated with holding on tightly to NPM while significant change occurs; there is a gap in terms of studying actual processes to understand these tensions.

This article develops a theoretical lens based on foundational elements of NPM that are particularly salient for examining the PIF in New Zealand. This is followed by a brief explanation of the context for the PIF. We explain our methods and then the discussion explores the implications of the PIF using the lens developed in section two and integrating interview findings. The article closes with conclusions and suggested research going forward.

**From Performance Information to Performance Regimes**

This section considers how the focus on performance in the public sector has shifted from piecemeal approaches to gathering information about performance to multiple performance measures meeting a range of purposes.

Performance is a multi-dimensional construct with varying significance to stakeholders (Bovaird 1996, Walker and Boyne 2006). Depending on the stakeholders, performance information is used in different ways – passively, politically, perversely and purposefully (Moynihan 2010). Gathering performance information in the public sector has involved adaptation of performance measurement tools such as TQM and Balanced Scorecard. Judgements about the performance of public sector organizations may take the form of a ‘regime’, involving multiple instruments and even multiple institutions, all operating within systems of formal and informal rules and cultures (Martin et al. 2016). The normative assumption underlying performance management is that positive change will result from the activity of measuring and analysing (Radin 2006).

Increasingly, the effectiveness and sustainability of the performance of public sector organizations is not in a single organization’s hands but is co-created in inter-organizational networks (Osborne et al. 2015). This contemporary governance environment raises this problem of interorganizational dependency starkly. Frederickson and Frederickson (2006) suggested that the doctrine of performance management was based on a model of organizational improvement that involved few, direct providers in simple relationships with a bureaucratic state. The movement away from traditional bureaucracies and towards interorganizational networks creates accountability difficulties related to the measurement of performance. Adaptation in the performance regime is required, that takes account of behavioural responses to that performance regime – including the acknowledgement that behavioural responses might not be the intended ones. Where adaptation is slow or non-existent, the performance approach may be experienced as externally imposed standards that are unlikely to achieve desired change, resulting in passivity, gaming and possibly evasion (Jakobsen et al. 2017).

There are varying processes at work within what have been termed ‘performance regimes’. Archetypes provide ways in which to consider how performance management is undertaken, but also allude to progression or layering of approaches. Bouckaert and Halligan (2008) suggest four categorisations or archetypes: 1) performance administration; 2) management of performances; 3) performance management; and 4) performance governance. Moving through these approaches implies an increasing sophistication from information gathering to system monitoring, the performance system must be adaptable and the context appropriate. Depending on this context, it may not be desirable or even possible to move between the types.

**NPM effects and context**

The context in which we undertake our analysis begins with the assumption that NPM is persisting while complexity increases (Diefenbach 2009) Governance networks and newer forms of governance such as social investment and investing for well-being approaches involving a high level of collaboration across organizational units require more developed forms of performance management, but the simplicity of how most performance regimes are constructed is at odds with the increasingly inter-connected work directed towards achieving outcomes (Moynihan et al. 2011). Improving the outcomes associated with child poverty or high prison incarceration rates, for example, requires multi-organizational, multi-disciplinary teams, across sectors, at all levels, with budgetary flexibility. To some extent, the performance literature has set up expectations that ‘performance management can be a panacea for these thorny problems’ (Radin 2006, p. 234).

Efforts to overcome the organisational silo effects created by NPM reforms do not replace these earlier reforms but have added to them to create so-called hybrid post-NPM governance forms (Laegried and Christensen 2001, 2007a, 2007b, Christensen and Lægreid 2012, 2015; Lodge and Gill 2011, Xialong and Christensen 2019). Where partnerships and now ‘collective impact’ forms of organization have taken root NPM’s focus on performance efficiency remains. Pollitt (2016) even argues that pre-NPM type reforms associated with savings inputs – salary freezes, across the board percentage cuts – have re-emerged.

In the next section we further develop our frame by outlining three dimensions of NPM-inspired performance approaches that are the focus of the empirical work presented later in the article – gaming, synecdoche, and accountability.

**Gaming and Synecdoche**

The legacy and continuing influence of NPM (Marchand and Brunet 2017, Nitzl et al. 2019, Pollitt 2016) affects attempts to alter performance management approaches. We draw attention to the implications of this using the case of New Zealand’s PIF.

This section engages a lens based on three points: first synecdoche, or how ‘the part’ represents ‘the whole’ and how the ‘whole’ represents ‘the part’, is an underlying principle that aids in understanding how performance management actually works (Bevan and Hood 2006, Drew et al. 2017, Pollitt 2013) Second, while gaming is a classic human and organizational response to performance activities it is not necessarily or always ‘bad behaviour’ (Radin 2006). And third, the conflict between performance management as ‘control’ and the nature of necessary accountability is problematic (Lewis 2015).

Many performance systems use a whole range of indicators that are typically reviewed individually and then represented as a ‘dashboard’. The ‘syncedochical gap’ (Drew and Gamage 2019) between what individual or grouped pieces of information mean and how that relates to the indicators taken in their totality is significant. Different ways of gathering information through scoreboards or functional assessments only tell part of the story. Because the full narrative of the performance of an organisation, or a true assessment is impossible to tell through indicators, synecdoche occurs whereby a subjective interpretation of the numbers and indicators becomes ‘the’ story and then depending on the nature of this story it may drive gaming behaviour. Relatedly, the full narrative may pose as ‘truth’ vis-à-vis individual indicators which taken individually or in groups would make meaning (Drew et al. 2018). Gerrish (2016) reports on a large literature and finds dysfunctional responses in performance systems using targets. While these effects may be unintended consequences (Franco-Santos and Otley 2018), they are well known and even accepted as part of the process. In line with Drew et al. (2018) we do not view gaming entirely as a deceitful practice aimed at wilfully massaging measurements in order to improve an individual or organisation’s position, but more as a set of behaviours or intentions that aim to change perceptions when viewing that organisation.

If gaming is the expected behavioural reaction to performance management efforts, where improvement is the explicit purpose, and control is the implicit purpose (Lewis 2015) then performance management design must consider its inevitable influence. Gaming results when the measuring mechanisms do not align with the intended objectives of the work and the intrinsic motivations of the individuals giving their account (Busuioc and Lodge 2016). People change what they do as soon as performance measures are in place, if they are aware that the data collected will be used in some way that affects their work, particularly in controlling how they work, they will change their behaviour (Lewis 2015). The ‘performance paradox’ or distance between performance on paper and actual performance may be caused by measurement errors or unintended distortions, the synechochical gap or intended distortion such as cheating or gaming (Drew and Gamage 2019). Gaming suggests consciously bad behaviour, but this is not always the case – it may be a means for meeting required targets in difficult circumstances and a rational move where principals may prefer regime control over performance (Jakobsen et al. 2017).

The design of a performance management system also raises questions about what it is for, who it is for, and to what extent it is really about control (Lewis 2015, Radin 2006). There is an assumed rationality underlying performance management - that some form of review or control will lead to a desired change. If performance measurement is ‘for’ improvement, then there is an assumed starting point from which ‘good’ change leading to improvement can come. Clear change arising from management actions and linked directly to measurement is an assumption at the root of performance improvement. Yet causality is far from clear with conflicting, ambiguous and complex public sector goals combined with the paradoxes of outcome indicators (Jackson 2011, Lowe 2013). The pressure to ‘comply’ to fulfil the clear goals on the one hand while being asked to think longer term about improvements and strategy is likely to lead to cognitive dissonance – and thus, gaming.

We now turn the third element framing our analysis, that of internal and external accountability and issues arising therein.

**Internal and external accountability**

First, accountability remains a fundamental element of performance management systems and the ways in which this is embedded in the design links very strongly to the influences of NPM. Second, reputation is a strong driver in terms of the ways in which officials respond to performance management. Third, the publicness of performance management affects how officials prepare for, experience and follow-up from a performance review.

Accountability in an NPM system is achieved through performance standards based on the managerial and normative assumptions that the act of examining performance will result in beneficial change and even improvement. Organizational disaggregation was a key element of classical NPM (Dunleavy 1994). The question of accountability became increasingly acute with the creation of substantial numbers of arms-length agencies. The answer to this was accountability through performance standards often uniformly applied. By focusing the design of performance measurement on outcomes and outputs and what is being achieved, the resulting transparency and financial accountability were imagined to also mean improvement (Hoque and Adams 2011). Yet these relationships have not been well-evidenced. In a meta-analysis, Gerrish (2016) looks for the association between performance management and performance. Significant value trade-offs are often made, and he suggests that possibly the focus should be on alternatives to performance management such as developing a public service motivation or ethic among managers.

The traditional explanations of behaviour associated with accountability are underpinned in most accounts by principal-agency. Busuioc and Lodge (2016) purport that the resulting accountability-enhancing measures lead to a host of negative behaviours such as cheating and gaming and even a reduction of intrinsic motivation and moral responsibility. However, they also argue that there is another approach that explains the actions of stakeholders. Principal-agency explains only part of what is going on among stakeholders whereby many individuals and organizations have an impact on what any one organization does or can do. Hybridity in governance as a response to this complexity is posed as one adaptation, but the underlying behavioural inclinations of public managers remain rooted in reputation-holding (if not enhancing) activity, as well as blame-avoidance (Drew and Grant 2017). Busuioc and Lodge (2016) suggest that, ‘Reputation holds the key to accounting for selective attention and variations in degrees of interest, intensity and investment in accountability processes… the way actors present themselves to, and are perceived and assessed by, a wider set of audiences matters’ (p. 248). Being under the pressure of targeted measurement that is scaled up to dashboard reporting in a publicly reported review, organizational and personal reputation management becomes a significant factor.

Individuals and organisations are being compared amongst themselves and are benchmarked at least implicitly if not explicitly. From a principal-agency point of view, in democratic systems the public also plays the role of principal in the interaction with politicians but with respect to performance management rarely does the public have a specific influence on organizational change at a detailed level (Mizrahi 2017). But in terms of reputation, the public plays an influential role. The conscious decision where results are available internally as well as online for public scrutiny has specific kinds of impacts on the organizations and individuals being performance reviewed. Government/s could choose to not make performance management results public, and indeed not all are (see Canadian MAF in 2009, Lindquist 2009). The decision to do so, to achieve accountability in the public domain is linked to the notion that external accountability will increase performance (Jakobsen et al. 2017).

**The New Zealand context**

This article focuses on the performance management landscape in New Zealand since 2009, when the Performance Improvement Framework (PIF) was launched. In this same period, the Better Public Services (BPS) reform programme commenced after a decade of concerns about the fragmentation caused by the departmental structural silos created by NPM. Of concern was the lack of focus on results combined with an increased need for efficiencies influenced by the Global Financial Crisis.

The label PIF is applied to both the conceptual framework for evaluating public sector organizational capability and performance and the operational process through which a four-year future focus on organizational capability and performance is assessed.

Originally based on the UK’s Capability Reviews (Te Kawa and Guerin 2012), the framework was adapted to the New Zealand context and has been upgraded four times in response to evaluation and reflecting the changing public sector management focus on targets as represented in the BPS programme. Of note is the ‘Four-year Excellence Horizon’ (introduced from 2011), and the change in 2015 to a focus on customers and the delivery of value to New Zealanders. The PIF is described by the agency responsible for its operation, the State Services Commission (SSC) as a *‘uniform, yet flexible, framework to support continuous performance improvement across the State Services’*(SSC n.d). Every public service department and most major Crown entities have had PIF Reviews and many have had Follow-up Reviews.

A parallel reform programme, BPS, put more emphasis on achieving better outcomes for New Zealanders as a whole. The Government established a ‘to do’ list of 11 high priority goals or result challenges with measurable targets (DPM and Minister of State Services 2012b, Scott and Bardach 2018). The targets included, for example, specific increases in childhood immunisations and decreases in long-term benefit liabilities; and relied heavily on a new approach to data analytics encapsulated in the ‘social investment approach’ (Boston & Gill 2017).

The enabling legislation and implementation of these reforms since 2012 shifted New Zealand’s public management model to a more complex hybrid one where achieving results based on high-levels goals required the collaboration of multiple public sector agencies and stewardship for the future (Scott and Boyd 2016, 2017).

In January of 2018 the new Labour-led government cancelled the BPS and changed the way in which it would measure with both new measures of (absolute) child poverty, new legislation, and different kinds of explicit targets such as reducing the number of cold damp homes instead of proxies such as reductions in rheumatic fever, or numbers gaining school qualifications. While the BPS was rejected, the PIF was retained and remains an active performance management programme. A single organization perspective on performance continued largely in its 2016 version with emphasis on providing an organizational narrative combining operational performance reporting with strategic expectations.

**Methodology**

In 2016/17, an evaluation of the PIF was undertaken by a team at the School of Government, Victoria University and was published by SSC in 2017(Allen et al., 2017). This involved a multi-method approach: a) historical and contextual background including an international review of selected comparator systems, b) a qualitative study based on the experiences of the PIF in three agencies, and c) a quantitative survey responded to by 430 tier one, two and three public servants. More in-depth perspectives and experiences of the PIF were obtained from the qualitative study involving 27 participants in mainly senior roles in the three case study agencies. The survey reported on, for example, the extent to which the PIF produced useful information for the organization about its strategic framework, whether the quality of the external review and capability of the lead reviewers was high, and whether benefits of the PIF process were worthwhile.

This article focuses on the qualitative results from the interviews with knowledge of the historical study and the qualitative comments that were added to survey returns. As well as the empirical data the research team had access to and used a range of government documents associated with the conduct of the PIF reviews back to 2009. This documentation provided an in-depth understanding of the development and evolution of the PIF and ensured that our study involved consideration of a wide range of organizations and PIF experiences.

A process of getting to agreement on the agencies to be studied was undertaken. The agencies agreed upon were not extreme cases in terms of positive or negative PIFs but exemplified a range of challenges in the context of New Zealand organizational performance. The final list was reduced to three organizations representing a traditional Ministry, a Crown entity and a public service Department. While the sample was small it was a good representation of institutional form and service delivery in New Zealand and interviewees represented a cross-section of senior level leadership in the organizations (Denzin and Lincoln 1994). The twenty-seven interviews provided a substantial set of data on the operation of the PIF. Saturation and replication was evident in the interviewing process (Morse 2002).

Interviews took place with reviewers, senior executives, middle managers and central agency senior officials, as well as with a limited number of external stakeholders. Interview data were coded using thematic networks, working through the stages of coding material, identifying themes, constructing thematic networks, exploring the networks and interpreting patterns (Attride-Stirling 2001). Two team members undertook the coding separately and themes were discussed iteratively over consecutive meetings. The thematic coding frame was shared with the research team and tested with two other team members for rigour.

Ethical approval for the project was provided by the University of Victoria’s Research Ethics Committee (0000022776). Interviews were conducted face-to-face, with one conducted by skype. Interviews were semi-structured, recorded with consent and anonymized. Quotes used in this paper have a series of identifier codes due to reduce the potential for individual or organizational identification. The data was used to uncover emergent themes about the operation and meaning of the PIF, from those who had experienced being involved in one or more.

**Investigating the PIF**

The PIF is a product of the New Zealand NPM-inspired public management arrangement that evolved alongside wider experiments with governance such as the BPS program. We analyse our interview data based on our theoretical lens focussing on the effects of gaming, synecdoche, and accountability. The stages of the PIF provide the backdrop against which we consider the interplay of these effects.

Figure 1 here

The first stage of the PIF involves a decision by SSC to review an organization based on a rolling cycle or a problematic organisational issue. Several actions are triggered including in-organization gathering of all performance information, reviewing past performance resulting in a Self-Review, and a process of negotiation over timing and which Lead Reviewers will be assigned to the organization. One senior official remarked: *we had to* *make sure all the strategic leadership team and those that were going to be interviewed were fully conversant with the types of issues and things we were facing and things we were going to make sure that the PIF reviewers knew (SR 4).* This can be interpreted as making efforts to ‘getting our story straight’ and shaping the performance information in such a way as to shine the best possible light on the organisation (Van de Walle and Roberts 2008). Managing the ‘performance paradox’ - the discord between the performance on paper and the actual performance (Drew and Gamage 2019) – continues throughout the PIF.

The efficacy of the PIF revolves around the Lead Reviewers – highly experienced individuals with deep functional knowledge of New Zealand public sector organisations as well as a strategic orientation. Getting the ‘right’ Lead Reviewers became an increasingly important issue over time, particularly after SSC’s publication of findings about system performance in 2013 (SSC et al. 2013). Chief Executives are consulted on to who is going to review their organisation. The choosing of Lead Reviewers (usually two) at the outset of the PIF process involves the possibility of both intended and unintended distortions (Drew and Gamage 2019). A senior departmental official remarked,

*And we took some time to ensure that we got a reviewer … that was going to challenge but that understood enough about our environment without necessarily having too many preconceived ideas – [X] was perfect. (CA3b)*

There was opportunity for organisations to obtain a sympathetic Reviewer, as well as a Reviewer that was suited to the style of the CE and the type of challenging conversations the CE was hoping to have.

*There are certainly some CEs who chose lead reviewers, and I think they knew they were choosing lead reviewers that wouldn’t be as challenging as other lead reviewers.  I also saw agencies deliberately choose lead reviewers who would lift their thinking. (SM4)*

While gaming is undoubtedly a feature of this process, the behaviour is not cheating (Drew and Gamage 2019). The process is a rational response to the way in which the PIF is constructed and the lack of control organisations have in complicated problem areas involving many stakeholders. Negotiating who reviews the organisation, and when, is a critical role for the CE. The PIF is both a control mechanism (Lewis 2015) as well having other legitimate goals of aiming to align the organisations’ goals with an as yet-undetermined future state. This is an exceedingly challenging situation that is one result of efforts to adjust a performance management system to benefit from dialogue and altered dynamics (De Bruijn 2007)

The in-organisation review (PIF second stage) involves Lead Reviewers interviewing individuals throughout the organisation and external stakeholders. The Reviewers have already examined the organisation’s pre-PIF preparations around functional and operational performance as well as any previous PIF reviews. Through rigorous questioning, the interviewing process validates this information, interrogates it further and involves dialogue with interviewees to extract and understand how the organisation connects to its stated goals.

The in-organization process, after an intense period of preparation, can be a highly pressurised experience in an environment of complex accountability requirements.

*We live on taxpayers money so I get that: the public nature, but in any one year what happens? We get audited, we go to Select Committee twice, we have Ministers, we’re in Cabinet Committees the whole time on various things, with our budget round we’re in front of Treasury, we’re in front of the Minister of Finance – we spend a hell of a lot of time getting QA’d. Then here comes the PIF, and you really feel like it’s survivorship mode by the time the PIF comes along. (FR21)*

While recognising the potential benefits of the PIF, officials felt the weight of a resource intensive process. The design of the PIF has shifted in the attempt to create a more holistic and strategic form of performance review, toward ‘improvement’ on a four-year horizon, but there is a significant tension arising from whether it is viewed as a compliance exercise or an improvement exercise or something other than that.

*What I’m saying is, that there would lessons learned from it, and all the rest of it, but I would approach it as there’s going to be a huge compliance burden, and there’s going to be a really big risk management exercise here. (JO5)*

This tension emerges even more clearly where officials voice their feelings about how despite efforts to move the PIF into a more supportive, strategic mode, it remains or retrenches into a compliance or assessment exercise. One official seems to indicate that they are okay with this but would prefer clarity rather than a weak effort to ‘assist’.

… no matter how strong the assurance that we’re doing ‘assist’, there’s always the back of the mind saying, ‘nah, this is going to end up in the ‘assess’. So I think having that independence [of the PIF reviewers] and just making it very clearly articulated about what this is – that works. I think it’s really hard to be the assess, and the assist. (FS18)

After the review, in the third stage, the Reviewers write up a structured detailed report. The report has two parts, a benchmarked functional review combined with a more narrative-based strategic review. A dashboard is presented with ‘green, yellow and red-light’ (or hashed combinations) indicating which areas are identified as needing improvement or are failing. The problem of synecdochical gap (Drew and Gamage 2019) is potentially addressed here: where the dashboard tells one part of the story (the measured elements easily interpreted as the whole story, ‘part for the whole’), the interviewing, dialogue and qualitative aspects address elements of performance that cannot possibly be represented in numbers or graded colours (but could be interpreted as ‘the whole for the part’). The problem is that the colours and dashboard approach focus viewer attention and the long narrative about the organisation will only be addressed by the senior people in the organisation. If what is represented in the narrative is not absorbed by senior management and conveyed effectively in terms of relating present performance to future expectations, the only thing left is *the effect* having got ‘red’ or ‘yellow’ will have on organisational members.

The ratings are problematic and an ongoing design issue for the PIF.

*We toss up all the time whether the ratings are the right tool…There are some organisations that would read the four-year excellence horizons with no ratings and ignore it and get on. But the ratings do focus the mind for some reviews but for some reviews it does get in the way, it’s demoralizing for staff, it doesn’t help the public perception (if the ratings are bad), there is quite a tension there.* *(CV15c)*

Over the years, organisations have learned ‘how’ to do a PIF and the assessment approach supports this kind of behaviour; there is benefit in watching and studying how the organisations who earn the best scores do it. *If you really want people to learn from it and use it that way then you can’t score them a mark out of ten – they’ll just find out which agencies are best at managing the process. (FT129)* The likelihood of ‘data distortion’ is significant given the pressure to show improvement over time (Drew and Grant 2016)

Despite the combined approach with additional layers of reporting around strategic issues and future-looking ambitions, the nature of compliance related to the PIF remains a vexing issue. The additional requirements have not meant that the scoring disappears, if anything the dashboards have become the only meaningful way to interpret the PIF reviews because of the more abstract nature of the strategic requirements and the confusion inherent in interpreting the final PIF reports. The synecdochial risk of the ‘whole’ representing every part, may mean the details of the operational issues get lost. CEs are essentially the only people in the organisation who can connect the full extent of the operation with the future looking requirements thus rendering the strategic exercise moot for the majority of the organisation.

How can ratings sit comfortably with judging organisations on their ability to strategize over a four-year (or longer) time horizon? Perhaps they cannot – this NPM-style performance reporting drags strongly towards a target focus (Bevan and Hood 2006), and now we add to this accountability issues and related concerns over reputation.

In the fourth stage the draft report written by the Lead Reviewers is discussed, and negotiated; organisations receive their report and then make decisions as to when or if they act on the recommendations. The draft is viewed by the organisation’s CE and conversations ensue between Lead Reviewers, SSC, Department of the Prime Minister and Cabinet (DPMC), Treasury and the CE on the implications of the report. The CE has a substantial vested interest in how the report is written and what it says about the organisation.

*The other part I think that starts to occur is that you recognize your organisation is going to be publicly named and in the spotlight in terms of what you are and aren’t doing…. So you start to put that risk mitigation and political lens across it. If there’s a tension around some of the story, you’re seeking to have that discussion with reviewers to say well, we get you but we wouldn’t put it that way for these reasons. And it’s the reviewers right to say, well, we would, or fair call. (CB10c)*

CE’s are placed in a position of responding to efforts to influence the organisation by SSC and the Treasury with respect to its functional performance, as well as having to think about the longer-term strategic role of the organisation in terms of serving New Zealand citizens. While the discussion of an organization’s report among SSC, the Treasury and DPMC is designed to create an understanding of current performance against future challenges and interdependencies there was little to evidence that this part of the process was effective.

The PIF becomes a major exercise in reputation management – managing and cultivating both an organisation’s and individual’s reputation vis-à-vis different audiences (Carpenter 2010). In a centralized yet disaggregated system with CE’s as major players, organisations are concerned about how their reports rate against other comparable organisations and how their Minister and other stakeholders will be affected by the review. PIFs are not formally ranked (though internally benchmarked) but informally everyone knows about a poor PIF and how the central agencies view the CEs. The CE finds she or he must convey the impression of running a competent organisation to wide networks of audiences that may only be ‘loosely connected to the formal political-institutional accountability mechanisms that principal-agent is so fond of’ (Maor and Sulitzeanu-Kenan 2016).

The accountability requirements go beyond pure reporting and to a large extent the PIF’d organisations see managing their reputations within the public sector and beyond with stakeholders as a critical part of the process and related to the fact that ‘everyone sees’ their results.

*I don’t mind the wise counsel and all the rest of it.  The tricky thing is you get someone and they PIF you, it’s published. You spend huge amounts of time working with stakeholders, you’ve got high levels of anxiety about are they or aren’t they going to endorse what you do? (SY 1)*

Although interest outside of government circles has been limited in terms of public exposure and the PIF has largely gone under the public’s radar, CEs and the organisations are concerned about their reputation and the ways in which they will be viewed by their stakeholders. The mere fact that a poor report could be picked up by the media has an impact. W*hat’s going to come out of this report - how am I going to manage it with ministers - how am I going to manage it with stakeholders? What happens if it ends up on the front pages of the [paper].* *(MB5)*

The misalignment of NPM-influenced performance management processes (Nitzl et al 2019) holds back the potential for the dialogue related elements; and public managers end up focussing on the reputation-protecting or reputation-enhancing elements that can help them find some sort of balance in face of the complexities of running their organisations (Busuioc and Lodge 2016). An organisation’s public reputation is known to have a positive impact on the organisation’s performance (Lee and Whitford 2013), consequently behaviour will be directed towards maintaining it.

While CE’s are being asked throughout the PIF process to consider the future and how the organisation can meet the needs of New Zealand, the political focus remains short-term. Confusion emerges with the inevitable contradictions embedded in a measuring process that has a future-looking strategic intent in a system with a 3-year election cycle:

*So we had a very difficult conversation with Minister X when she was Minister for [] on Lead Reviewer Y’s review of the [Department], and she was saying, I don’t want to care about a four year time horizon - I want to get things done in four months - that’s my focus. Why is the CE worrying about this four year stuff? (FR69)*

To some extent the PIF tries to deal with this through the experience and expertise of the reviewers:

*There is a designed ambiguity in the PIF, in that the assessment is against what you need to be to get your sort of excellence horizon in four years time. So, by definition almost, inevitably you’re not ready yet. People struggle with that; we’ve done all this work - look how far we’ve come, and I’m still a red/orange . … So there’s actually a bit of a tension in that. That’s particularly so in the assessment space, I think because it’s a report card against a future scenario. (FT56)*

It is an ambitious expectation to combine reporting on past performance, assessment of current operations and strategic thinking for four years ahead into one performance management tool.

The inclusion of a narrative exploration of the organisation, based on how well it is poised for future challenges may partially balance the effects of gaming given its emphasis on reflection on the strengths and weaknesses of the organisation against an unknown scenario. Adaptation in this performance regime has potentially improved the possibilities of what a performance system can do. However, the other NPM effects constrain the PIF in terms of how it might be used to better effect in the evolving NZ system.

**Conclusion**

The New Zealand PIF is an example of a type of performance management system that is aiming to influence how organisations change and improve over time. This has been designed as a combination functional reporting with forward-looking strategic consideration of what an organisation needs to do into the future to meet the needs of New Zealanders. The tensions arising from gaming, distortions (intended and unintended consequences) and the effects of reputation-protecting behaviour are partially offset by the introduction of dialogue during the preparation phase, dialogue and learning during the reporting phase and the important role of outside Lead Reviewers.

The Lead Reviewer approach and its interviewing format, introduces dynamics into a performance system that mitigates somewhat against the impact of the ratings approach, as does the way in which a PIF review is presented and published. Yet, gaming behaviour persists, and the common graded ratings continue to hold the attention of those to whom the PIF is important, acting as a proxy for the whole performance of the organisation. Where the organisational narrative holds importance, notably to the central agencies, the internal performance successes may be overlooked. The publicness of the PIF and the accountability arrangements through the SSC but also to Ministers, stakeholders and internal employees ensure transparency but influence reputation-protecting behaviour ensures the PIF is seen as a risk management and compliance exercise.

By theorising about the PIF in this way, we have provided a case study of how, despite adaptations, a performance management system is limited by the ubiquitous effects of NPM. The enhanced elements that are meant to be pushing the system to an ‘improvement’ focus, are in tension with the NPM-style measurements and raise further questions as to the usefulness of resource-intensive performance management systems (De Bruijn 2009). Additional research is required to address the complexity of the New Zealand performance management system to determine whether the moves towards hybridity in governance require a revamped or different system that can address the needs of the public sector and the New Zealand public. Further learning can be captured from examining the extent to which the experience of the PIF in New Zealand can contribute to other comparative systems.

References:

Allen B., Berman E., Cantal C., Eppel E., Jackson B.., Lofgren K, Macaulay M., Phillips B. and Plimmer G. 2017. Independent Review of The Performance Improvement Framework. Victoria School of Government, Wellington, New Zealand. i-45

Attride-Stirling, J. 2001. ‘Thematic networks: an analytic tool for qualitative research’. *Qualitative Research* 1(3): 385-405.

Bevan, G. and Hood, C. 2006. ‘What’s Measured Is What Matters: Targets and Gaming in the English Public Health Care System. P*ublic Administration,* 84: 517-538.

Boston, J. and Gill, D. (Eds.) 2017. *Social investment: A New Zealand policy experiment*. Wellington: Bridget Williams.

Bouckaert, G., and Halligan, J. 2008. *Managing performance: International comparisons*. Abingdon: Routledge.

Boyne, G.A. 2010. ‘Performance management: does it work?’ In *Public Management and Performance: Research Directions*, eds. R.M. Walker, G.A. Boyne, and G.A. Brewer. Cambridge: Cambridge University Press, 207-226.

Bovaird, T. 1996. ‘The Political Economy of Performance Measurement’. In *Organizational Performance and Measurement in the Public Sector: Toward Service, Effort and Accomplishment Reporting*. Westport, CT: Quorum Books, 145-161.

Busuioc, E. and Lodge, M. 2016. ‘The Reputational Basis of Public Accountability’. *Governance: An International Journal of Policy, Administration and Institutions*, 29(2): 247-263. and Lodge 2016

Carpenter. D. 2010. *Reputation and Power*. Princeton, NJ: Princeton University Press.

Christensen, T. 2011. ‘Complexity and Hybrid Public Administration – Theoretical and Empirical Challenges’. *Public Organization Review* 11(4): 407-423.

Christensen, T., & Lægreid, P. 2012. `Governance and administrative reforms`, ed. D. Levi-Faur, *Oxford handbook of governance*. Oxford: OUP, 255-267.

Christensen, T. 2012. ‘Post-NPM and changing public governance’, *Meiji Journal of Political Science and Economics*, (1): 1-11.

Christensen, T. and Lægreid, P. 2015. `Performance and Accountability – A Theoretical Discussion and an Empirical Assessment`. *Public Organisation Review*, 15: 207-225.

De Bruijn, H. 2009. Performance, performance, everywhere… ‘Critical books review’, *International Review of Administrative Sciences*, 75(1): 206-211.

De Bruijn, H. and van Helden, J. 2006 ‘A Plea for Dialogue Driven Performance-Based Management Systems: Evidence from the Dutch Public Sector. Financial Accountability & Management 22(4): 0267-4424.

Deputy Prime Minister (DPM) and Minister of State Services 2012b. *Cabinet Paper - Better Public Services Results: Targets and public communication*. Available from: http://www.ssc.govt.nz/bps-cab-papers-minutes.

Denzin, N. K., & Lincoln, Y. S. (Eds.) 2011. *Sage handbook of qualitative research* (4th ed.).Thousand Oaks: Sage.

Diefenbach, T. 2009. `New Public Management in Public Sector Organisations: The Dark Sides of Managerialistic ‘Enlightenment’`, *Public Administration*, 87(4), 892-909.

Drew, J. and Gamage, S. (2019) ‘Just do it? A cautionary tale on implementing performance management regimes’, in ed. Pilcher, R. and Gilchrist, D. *Public Sector Accounting, Accountability and Governance – Globalising the Experiences of Australia and New Zealand*. London: Routledge, 105-118.

Drew, J. and Grant, B. 2016. ‘Means, Motive and Opportunity – Local Government Data Distortion in a High-Stakes Environment’. *Australian Journal of Public Administration* 76(2): 237-250.

Drew, J. and Grant, B. 2018. Multiple agents, blame games and public policy-making: the case of local government reform in New South Wales. *Australian Journal of Political Science*, 52(1): 37-52.

Drew, J., O’Flynn, J., and Grant, B. 2018. ‘Performing What? Exploring and Expanding the Notion of Synecdoche in Performance Management Practice’. *Public Administration Quarterly* 42(3): 395-424.

Dunleavy, P. 1994. ‘The Globalization of Public Services Production: Can Government be ‘Best in World’?’. *Public Policy and Administration* 9(2): 36-64.

Frederickson, D. and Frederickson, H. 2006. *Measuring the Performance of the Hollow State*. DC: Georgetown University Press.

Franco-Santos, M. and Otley, D. 2018. ‘Reviewing and Theorizing the Unintended Consequences of Performance Management Systems’. *International Journal of Management Reviews*, 20(3): 696-730.

Gerrish, E. 2016. ‘The impact of performance management on performance in public organizations: A meta-analysis'. *Public Administration Review* 76: 48-66.

Halligan, J., Sarrico, C. S., and Rhodes, M. L. 2012. `On the road to performance governance in the public domain?` *International Journal of Productivity and Performance Management, 61*(3), 224-234.

Hoque, Z. and Adams, C. 2011. ‘The Rise and Use of Balanced Scorecard Measures in Australian Government Departments’. Financial Accountability & Management, 27(3): 308-334.

Jakobsen, M., Maekgaard, M., Moynihan, D., van Loon, N. 2017. ‘Making Sense of Performance Regimes: Rebalancing External Accountability and Internal Learning’.  *Perspectives on Public Management and Governance*. 1-15.

Jackson, P. 2011. ‘ Governance by numbers: What have we learned over the past 30 years?’ *Public Money & Management*. 31(1): 13-26.

Lee, S. and Whiftord, A. 2013. `Assessing the effects of organizational resources on public agency performance: Evidence from the US federal government`. *Journal of Public Administration Research and Theory*, 23, 687-712.

Lewis, J. 2015. ‘The Politics and Consequences of Performance Measurement’. *Policy and Society* 34(1): 1-12.

Lindquist, E.(2009) ‘How Ottawa Assesses Department/Agency Performance: Treasury Board’s Management Accountability Framework’. In How Ottawa Spends, 2009-2010 – Economic Upheaval and Political Dysfunction. Ed. A. Maslove. Montreal: McGill-Queen’s University Press:

Lodge, M. and Gill, D. 2011. ‘Toward a New Era of Administrative Reform? The Myth of Post-NPM in New Zealand. *Governance: An International of Policy, Administration, and Institutions*. 24(1): 141-166.

Lowe, T. 2013. ‘New development: The paradox of outcomes – the more we measure, the less we understand’. *Public Money & Management* 33(3): 213-216.

Maor, M. and Sulitzeanu-Kenan, R. 2016. ‘Responsive Change: Agency Output Response to Reputational Threats’. *Journal of Public Administration Research and Theory*, 26(1): 31-44.

Mayne, J. And Zapico-Goiii, E. (Eds.) 1997. *Monitoring performance in the Public Sector: Future directions from international experience*. London: Routledge.

Martin, S., Nutley, S., Downie, J. and Grace, C. 2016. `Analysing performance assessment in public services: How useul is the concept of a performance regime?`. *Public Administration,* 94(1): 129-145.

Mizrahi, S. 2017. *Public Policy and Performance Management in Democratic Systems, Theory and Practice*. Cham, Switzerland: Palgrave Macmillan.

Moynihan, D., Fernandez, S., Kim, S., LeRoux, K., Piotrowski, S., Wright, B. And Yang, K. 2011. `Performance regimes amidst governance complexity`. *Journal of Public Administration Research and Theory,* 41(Supplement 1), i114-i155.

Moynihan, D. and Pandey, S. 2010. ‘The Big Question for Performance Management: Why Managers Use Performance Information?’. *Journal of Public Administration and Research and Theory*, 20:849-866.

Moynihan, D. 2008. *The Dynamics of Performance Management: Constructing Information and Reform*. DC: Georgetown University Press.

Moynihan, D. P. 2010. From performance management to democratic governance. In R. O'Leary, D. M. Van Slyke, & S. Kim (Eds.), *Future of public administration around the world* (pp. 21-26): Georgetown University Press.

Morse, J., Barrett, M., Mayan, M., Olson, K., Spiers, J. 2002. ‘Verification Strategies for Establishing Reliability and Validity in Qualitative Research’. *International Journal of Qualitative Methods* 1(2): 13-22.

Nitzl, C., Sicilia, M., and Steccolini, I. 2019. ‘Exploring the links between different performance information uses, NPM cultural orientation, and organizational performance in the public sector’. *Public Management Review*, 21(5): 686-710.

Osborne, S. (Ed.) 2010. *The new public governance? Emerging perspectives on the theory and practice of public governance*. London: Routledge.

Osborne, S., Radnor, Z. and Nasi, G. 2013. `A new theory for public service management? Toward a (public) service-dominant approach`. *The American Review of Public Administration,* 43(2): 135-158.

Osborne, S., Radnor, Z., Kinder, T. and Vidal, I. 2015. ‘The Service Framework: A Public-Service-Dominant Approach to Sustainable Public Services’. *British Journal of Management*. 26(3): 424-438.

Pollitt, C. 2013. ‘The Logics of performance management’, *Evaluation* 19(4): 346-363.

Pollitt, C. and Bouckaert, G. 2011. *Public management reform: A Comparative Analysis – New Public Management, Governance and the Neo-Weberian State*. Oxford: Oxford University Press.

Pollitt, C. and Bouckaert, G. 2017. *Public management reform: Into the Age of Austerity*. 4th edn. Oxford: Oxford University Press.

Pollitt, C. 2018. ‘Performance management 40 years on: a review. Some key decisions and consequences’. *Public Money & Management*. 38(3): 167-174.

Pollitt, C. 2016. ‘Managerialism Redux?’ *Financial Accountability and Management*. 32(4): 429447.

Radin, B. 2006. *Challenging the performance movement: Accountability, complexity, and democratic values*. Washington D.C.: Georgetown University Press.

Rhodes, M., Biondi, L., Gomes, R., Melo, A., Ohemeng, F, Perez-Lopez, G., Rossi, A., and Sutiyono W. 2012. ‘Current state of public sector performance management in seven selected countries’, *International Journal of Productivity and Performance Management*, 61:3, 235-271.

Scott, R.. and Boyd, R. 2016. ‘Results, targets and measures to drive collaboration: Lessons from New Zealand Better Public Service reform’. In *The three sector solution: Delivering public policy in collaboration with not-for-profits and business* eds. J. Butcher and Gilchrist. Canberra: ANU Press, 235-257.

Scott, R. and Bardach, E. 2018. A comparison of management adaptations for joined-up government: Lessons from New Zealand. *Australian Journal of Public Administration*. 1-22. https://doi.org/10.1111/1467-8500.12348

State Services Commission, Treasury, & Department of Prime Minister and Cabinet 2013. *Core Guide 3: Getting to great; Lead Reviewer insights from the Performance Improvement Framework.* Wellington: State Services Commission. Available from http://[www.ssc.govt.nz/pif](file:///G:\Writing\Current\PIF\August%20version%20for%20submission\www.ssc.govt.nz\pif)

State Services Commission (n.d) Performance Improvement Framework. Available from <http://ssc.govt.nz/performance-improvement-framework>*.*

Talbot, C. 2010. *Theories of performance: Organisational and service improvement in the public domain*. Oxford: Oxford University Press.

Te Kawa, D. and Guerin, K. 2012. Provoking Debate and Learning Lessons – it is early days, but what does the Performance Improvement Framework challenge us to think about? *Policy Quarterly*. 8:4, 28-36.

Van de Walle, S. and Roberts, Al. 2008. ‘Publishing Performance Information: An Illusion of Control?’ In *Performance information in the public sector: How it is used*. eds. Van Dooren, W. and Van de Walle, New York: Palgrave Macmillan, 211-240.

Van Dooren, W. and Van de Walle, S. 2008. eds. *Performance information in the public sector: How it is used*. New York: Palgrave Macmillan.

Van Dooren, W., Bouckaert, G. and Halligan, J. 2015. *Performance management in the public sector* (2nd ed.). London: Routledge.

Walker, R. and Boyne, G. 2006. ‘Public Management Reform and Organisational Performance: An Empirical Assessment of the U.K. Labour Government’s Public Service Improvement Strategy’, *Journal of Policy Analysis and Management*, 25(2):371-393.

Xialong, T. and Christensen, T. 2019. ‘Beyond NPM to Post NPM? A Study of China’s Government Reforms over the Past 40 Years’, *American Review of Public Administration*, 1-11.