

# Reviewing “Measuring” and “Improving” State Sector Productivity:

*A review for the NZ Productivity Commission (NZPC) of the two final State Sector productivity documents produced by the NZPC in August 2018.*

Dr Murray Horn  
September 2018

## Introduction

This is a review, commissioned by the NZPC, of the two final reports it produced in response to the Minister of Finance’s request for an inquiry into how the NZ state sector can effectively measure and improve productivity in core public services, with a focus on health, education, justice and social support.

The terms of reference for this review are in Annex A and for the Commission’s inquiry (the “Terms”) in Annex B. The terms of reference for this review require that I assess the final reports against seven performance measures, which are labelled one through seven below. In assessing the quality of the Commissions work (performance measure number three), the Commission asked that I answer eight specific questions that assess different aspects of work quality (labelled 3.1 through 3.8 below).

This Introduction is followed by a summary assessment which is followed by a detailed assessment against each of the Commission’s seven performance measures. The two reports *Measuring state sector productivity* and *Improving state sector productivity* are referred to throughout this review as the *Measuring* and *Improving* documents respectively.

In undertaking this project I reviewed the NZPC’s documents, commissioned work and selected references and interviewed a number of stakeholders selected by the NZPC who were consulted by the NZPC during the inquiry. This included both the Treasury and the SSC who would be responsible for implementing most of the final recommendations accepted by Government.

## Summary Assessment

This summary assesses the *Measuring* and *Improving* documents against the seven performance measures required by NZPC. More detailed assessments on each follow this summary.

- i. **Right focus:** relevance and materiality in meeting the terms of reference

The Terms asked the Commission for advice on four areas in order to develop measures and recommendations to meet three objectives: provide Ministers with assurance about the performance of current delivery models; improve Chief Executives ability to understand and improve their business; and to engage workforces in opportunities to do things better.

The reports give relevant and material advice on most aspects of the four areas identified in the Terms. Where the inquiry falls short, it does so because the Terms created a dilemma for the inquiry: the best measures for meeting the three objectives in the four identified sectors are not the narrowly defined technical efficiency measures the Terms require.

ii. **Good process management:** the timeliness and quality of the inquiry process

The draft report was produced in a timely fashion and the final reports were delivered ahead of the deadline set out in the Terms.

NZPC adopted their usual process: i.e., inviting submissions to both an issues paper and the draft report, proactively seeking engagement meetings with interested and affected parties, commissioning research, engaging in roundtables, conferences and seminars and canvassing a wide range of published papers. This is a high quality process that elicited a good balance of views and evidence. However, the lack of proactive, Ministry-level engagement from most of the identified sectors would have compromised the quality of the process to some degree: i.e., affected parties did not seem all that interested. This is unlikely to have been caused by the way the Commission managed the process.

iii. **High quality work:** quality of analysis, guidance and recommendations

The guidance in the *Measuring* report was well pitched at a broad audience of potential users and is likely to be useful and used, if the feedback from interviewees is a good guide. On the other hand, for those just interested in applying narrow technical efficiency measures to operational activities, the comprehensive economic approach adopted is unnecessary.

The analysis identified the most important barriers to improving productivity in the state sector. The approach to improving productivity was pragmatic and realistic and the recommendations in the *Improving* report are practical, achievable and necessary (albeit possibly not sufficient) to address the barriers to productivity improvement identified.

There are a few things that I have identified as missing from the analysis and while adding these would have enhanced the final reports, the absence of these elements is not fatal to the analysis, guidance or recommendations in the reports.

iv. **Effective engagement:** how well NZPC engaged with interested parties?

Given the well-deserved reputation of the Commission for the quality of its engagement in past inquiries, I was surprised by the very wide assessments of interested parties in the quality of engagement in this inquiry. Some of that variation reflects different levels of unease about where the inquiry might lead and different levels of interest in both the subject (at least as specified in the Terms) and the process.

While these factors are largely outside the Commission's control there are some lessons for the Commission. First, some of the issues raised in the initial submissions could have usefully been given more attention. Second, concerns about the way the Commission might use or present what organisations are doing, or not doing, were not addressed (albeit that they may not have been addressable to the satisfaction of those organisations). Finally, narrow measures of productivity are likely to be much more relevant to staff managing operational activity and more could have been done to reflect more of an operational approach in the engagement (including having operational people with experience in using the narrow productivity metrics on the inquiry team).

v. **Clear delivery of messages:** how well is the work communicated and presented?

The work is well communicated and presented. This facilitates accessibility by a variety of different audiences. The only way this might have been improved is by adopting more of an operational approach and language in presenting the narrow technical efficiency measures and case studies (see iv above).

On the other hand, while the communication and presentation of “barriers to higher state sector productivity” demands attention, it does little to develop the confidence of the existing workforces that they will successfully overcome these barriers. This makes it less likely that the reports will meet one of the aims of the Terms: i.e., “to engage the State Sector workforces on opportunities to do things better.”

- vi. **Overall quality:** taking into account all factors

There are a number of factors outside the Commissions control that would have weighed on any inquiry. Despite that, the inquiry has produced an important and high quality contribution to our understanding of state sector productivity, its measurement and the barriers facing productivity improvement. It has provided useful guidance to officials who are interested in productivity measurement. Perhaps most significantly, it has identified a few practical, achievable and necessary recommendations that, if implemented, would go a long way to improving state sector productivity.

- vii. **Having intended impacts:** what happens as a result of the work?

While the *Measuring* report will be voluntarily picked up by a few officials keen on developing better measures, this is unlikely to be either a common or enduring response if the recommendations aimed at strengthening demand for productivity improvement are not adopted by government. Similarly, while Treasury and the SSC may well pick up some of the recommendations, unless Ministers support those measures aimed at increasing this demand, it is hard to see any new initiatives from central agencies having a significant and sustained impact. While it would be foolish to try and predict this Government’s response, the initial indicators are not encouraging.

## Detailed Assessments

### 1. **Right focus** – *the relevance and materiality of the final inquiry reports in meeting the terms of reference (the Terms)*

The Terms describe three objectives that addressing the current gap in good measures of productivity should aim to meet: provide Ministers with greater assurance about the “performance and innovation of current delivery models”; improve Chief Executives’ ability to “understand and improve their business”; and “engage the State Sector workforces on opportunities to do things better.”

In order to meet these objectives, the Terms requires advice on four specific areas:

- a. How to measure efficiency/productivity in health, education, justice and social support sectors (at sector as well as function or service level); including how imperfect measures are most appropriately and usefully employed.
- b. The appropriate role of measures in public sector performance frameworks, with the goal of improving assurance to Ministers and incentives on agencies for improvement.
- c. Developing the capability, culture and systems that can support agencies to better measure, understand and improve productivity.
- d. Provide advice on the best measures in the identified sectors.

The Terms makes it clear that the Inquiry “... should focus on the narrower definition of productivity as how efficiently inputs/resources are being utilised to generate quality outputs/services.” The Terms require the Commission to focus on technical efficiency (inputs required to produce each

output), not allocative efficiency (the right mix of outputs) or effectiveness (producing the desired result).

The Inquiry focusses on measurement at the service level; implicitly accepting that the approach to sector-level measurement that the Department of Statistics is developing. This makes sense because service-level measures best meet the “three objectives” described above: i.e., they are more relevant to Ministers, Chief Executives and in engaging the workforce.

The Inquiry also provides clear guidance on the “appropriate role of measures in public sector performance frameworks.” Recommendation R3.5 sets five conditions that should be met before productivity measures are used as “formal accountability measures.” Moreover, it explicitly recognises the tension amongst potential uses of productivity measures and is clear about where the balance should be: “While accountability and steering are important, the main benefit from productivity measurement is the potential to encourage conversations and learning about service improvement” (*Measurement* p. 10)

The Inquiry also makes small number of well-balanced recommendations that, taken together, would encourage agencies to develop the capability, culture and systems to better measure, understand and improve productivity as well as to support their efforts in doing so.

It is less obvious that the Inquiry has made “relevant and material” progress in suggesting “... how imperfect measures are most appropriately and usefully employed” and advising on “... the best measures in the identified sectors.” That is because it is extremely difficult to reconcile the direction to focus on “the narrower definition of productivity” with both the sectors and “three objectives” defined in the Terms.

The biggest potential gain from applying a narrow “technical efficiency” lens comes from focussing on services where the inputs consumed and the outputs delivered are easy to measure and where this narrow measure is more critical to success (typically operational activities where throughput is important to success). In this case the benefits of measurement are likely to be realised because these benefits are clearer to those directly involved in management of the service and easy for the workforce to accept as legitimate. Moreover the costs and risks of measurement are lower when measurement is easier. This is why we more often see these narrow measures routinely applied in operational activities, like call centres.

There are some relatively easy to measure operational activities in the named sectors, although concerns about the impact on increasing throughput on service quality are typically more important in those sectors than they would be in more operational departments, like IRD or Customs. It is not surprising, for example, that there is a good deal of interest in increasing hospital throughput (increased discharges and shorter lengths of stay). The “quality” impact (unplanned readmissions) is easy to measure and immediate, so relatively easy to manage. Moreover, the benefits of reduced wait times and increased elective procedures are valued highly by the public. On the other hand, increasing throughput in schools by increasing class sizes immediately raises concerns about quality that are harder to deal with: in part because there is less consensus about how to define quality; in part because the result (however measured) is typically co-produced; and in part because any impact on quality will take time to show through. So narrow measures of productivity in schools do not take you far without adding “quality adjusters” that move you much closer to what is of real interest, the effectiveness of increased educational spending (i.e., the broader measures of productivity ruled out of scope by the Terms).

In the sectors identified in the Terms, the effectiveness and the mix of services provided in delivering a desired result is typically far more important than the extent to which more of a particular service can be produced from a given quantum of labour and capital. While both aspects of productivity are important, Ministers and Chief Executives are far more likely to be focussed on the former than the latter in the named sectors. Moreover, measuring the service outputs and outcomes that really matter in the named sectors, and knowing what works best for different types of “client”, is typically harder and more controversial, so there is potentially a higher cost and higher risk associated with measurement. Finally, unless the measures are seen as reasonable by the workforce, it will be difficult to “engage” them in opportunities to improve measured performance.

This creates a dilemma: the best measures for the identified sectors in terms of meeting the three objectives for the Inquiry are unlikely to be the narrowly defined measures the Terms requires. While the Inquiry does try and make the narrow approach more relevant by adding assessments of the “quality” of service outputs, this can only go so far. On the other hand, just focussing on operational activity where the narrow definition would make most sense is unlikely to have provided Ministers with significantly more assurance about the performance of current delivery models or Chief Executives with significantly greater understanding of what they might do to improve their business, at least not in areas where the opportunities for assurance and improvement are likely to add most value.

While neither approach was likely to be particularly satisfying, the Inquiry could have enhanced its relevance in this area by identifying the factors typically cited that make measurement difficult and to suggest ways of addressing these problems (e.g., no counterfactual, attribution difficulties, uncertainty about what works, long delays in getting results). That may well have expanded the scope for the useful application of imperfect “quality adjusted” service measures by reducing the cost and the risk, and enhancing the likely benefits, of measurement.

Better to avoid creating these dilemmas by spending more time ensuring the inquiry is not over-specified in the commissioning phase. Once the aims of the inquiry are clear, as they were in this case, it would make sense to refine the inquiry Terms in consultation with the Commission. If ex ante specification risks over-specification, then better to leave room for Commission to use its judgement as the inquiry progresses. In this case, given the aim of helping Ministers, Chief Executives and the workforce with either assurance or improvement of the whole business in health, education, justice and social support, then better to let the Commission choose the productivity measure that best suits those situations. On the other hand, if focussing on technical efficiency really was central, then recasting the aim to focus on operational activities and giving the Commission scope to select sectors and organisations where this narrow measure made most sense would have worked better.

## **2. Good process management – *the timeliness and quality of the inquiry process***

**Timeliness:** The Terms required the Commission to publish a draft report and submit a final report to the Minister of Finance by 30 August 2018. The Commission received its Brief in June 2017, published an issues paper late July 2017, a draft report mid-December 2017 and final reports on 17 August 2018. Final Reports were delivered ahead of the deadline set out in the Terms.

**Quality:** The Commissions own process quality standard is “... to gather ideas, opinions, evidence and information to ensure that the inquiry is well-informed and relevant”. The Terms also required the Commission to consult with key interest groups and affected parties as well as “public sector agencies, those in receipt of public services and private sector agents that may have relevant insights.”

To this end the Commission:

- Sought submissions and held engagement meetings in response to both the Issues Paper and the Draft Report. It received 17 submissions and held 36 engagement meetings in response to the Issues Paper. The Draft Report elicited a further 14 submission (4 from those who had also submitted to the Issues Paper) and 10 more engagement meetings.
- It also commissioned 11 research papers, including a number of case studies.
- Engaged in two roundtables and four conferences and seminars.
- Referenced about ninety published papers.

Compared to the Low Emissions Economy Inquiry released shortly afterward, this attracted a very low number of responses from interested parties: e.g., 27 submitters versus 316 submitters for the Low Emissions Inquiry. Given the Commission adopts a similar process for its Inquiries, this limited response is most likely to reflect the degree to which affected parties were all that interested.

This may in part reflect the relatively narrow focus on the Inquiry (some people I talked to did reference this and it is interesting that IRD made a submission while most of Ministries from the sectors identified in the Terms did not). However, only 6 of the 27 submissions were from departments and none from the Ministries of Health, Education or Justice. While these Ministries did participate in engagement meetings (and some individuals from these Ministries were engaged in the process) this lack of proactive, Ministry-level engagement would have compromised the quality of the process to some degree.

While some people raised some concerns over the effectiveness of engagement (see discussion below) this does not explain the lack of proactive engagement from the relevant Ministries at the initial issues paper stage of the process. The most likely explanation for this lack of commitment and proactive engagement is that the same “barriers to lifting state sector productivity” identified on page 14 of the *Improving* Report would have been at work. This is hardly a reflection on the Commission’s management of the process.

In terms of the balance of submissions and engagement meetings taken together, there was a good coverage of various groups identified in the Terms.

Compared to the Low Emissions Inquiry, this Inquiry commissioned more research papers, presumably reflecting the relative scarcity of material available from the identified sectors (excluding health where the need was to scope what was being done) and need to illustrate the proposed approach with a few case studies.

Finally, there was a good attempt to incorporate international evidence and approaches. As well as the solid list of referenced material, eight of the engagement meetings were with international experts (from Australia, UK, US, OECD) with a mix of public, private and academic experts.

Overall, the Commission’s management of the process met the expectations set out in the Terms as well as the standards the Commission sets itself to ensure its Inquiries are well-informed and relevant. While the lack of proactive engagement and low level of commitment from some of the affected parties must have detracted, this is unlikely to have been caused by the way the Commission managed the Inquiry process.

### 3. High quality work – the quality of the analysis, guidance and recommendations in the final reports

The Commission seeks specific feedback on a number of issues directly related to each of the elements of this overall question. Addressing these questions in turn provides a comprehensive assessment of the question above, so I have decided to present my assessment under those headings.

This list of questions and comments follows.

#### 3.1 *Has the inquiry, in your view, advanced current thinking on the measurement of state sector productivity at a service level (Dunleavy (2015), Dunleavy & Carrera (2013), Van Dooren (2005)) by placing measurement within an institutional and performance management context?*

This is difficult to judge because there is not an agreed and comprehensive description of “current thinking” that is readily available.

The people I interviewed who had a view thought that the discussion on measurement gave a good description of the state of the art (or that the *Measurement* document brought together best practice). Some thought that the recommendations on the pre-conditions for using productivity as accountability measures were reasonably new, although similar conclusions have been drawn elsewhere.<sup>1</sup>

If we tested the proposition against what the Department of Statistics describes as the international consensus on “ideal” measures, then some tentative contributions have been made. For example, that output be measured from a consumer perspective so, in health, the whole course of treatment for a condition or disease is considered, rather than the individual activities that make up the treatment (so joined up primary and secondary care). The work commissioned as *Health Care Homes: Early Evidence in Wellington* was not available at the time the Final Reports were published but makes some small steps in this direction.

In terms of the cited works, Dunleavy (2015) suggests “...not to get into things like outcomes” (p28) and focuses on productivity in relatively easy to measure operational activities of operational departments (Customs, Tax and Benefit Administration). The Reports do go beyond this self-imposed constraint in assessing output “quality”, which only makes sense if the assessment of quality is related to the outcomes desired by consumers or the wider community.

Although this is not particularly new thinking, the way the Inquiry team tackled the subject did have elements that were new to me. In particular, the scoping elements of the nine step measurement method proposed by the Commission (summarised in table 2.1 of the *Measurement* report) does allow for measurement to be embedded in a broader “institutional and performance management context”. This is potentially new and interesting if this context is framed by the purpose of the organisation and the outcomes that it is created to deliver (much like the lens applied by the SSC’s PIF Reviews).

However, if this is the intention, it has to be read into the text because it is not spelt out clearly. While the business case includes “what the organisation might gain from measurement” and starts

---

<sup>1</sup> See, for example Cullen et al *Increasing productivity, reducing cost and increasing quality in elective care in New Zealand*, Internal Journal of Medicine 2012 42 (620-26) and the editorial by D Gorman and myself in the same issue.

with defining the business need the measure will address, this is cast in terms of learning versus accountability, rather than in terms of organisational purpose and business strategy. Similarly, while one of the guiding principles is that productivity measures complement measures of outcomes; this does not necessarily imply that these measures should be derived, or drawn, from the outcomes that the organisation is seeking to deliver.

Moreover, the two case studies in the Appendix that are meant to illustrate the application of the method do not apply the scoping elements to define the “institutional and performance management context”, at least not in a way that adds much value.

In short, while the “institutional and performance management context” has real potential to advance current thinking on the measurement of state sector productivity, the description of the method and its application in the case studies does not draw this potential out, at least in my view.

### *3.2 Did we pitch “Measuring state sector productivity” at the right level for its intended audience? Is it simple and comprehensive enough to guide productivity measurement in the state sector?*

This report is well pitched at a broad audience of potential intended users: i.e. “.. individuals and teams within the state sector who are intending to develop productivity measures” (see also answer to question 5). It is not too long or too technical (and technical terms are well described in the glossary).

The nine steps to executing the recommended technique are clearly spelt out (see table 1.2), although what is exactly intended by the scoping elements of the method are not totally clear (see 3.1 above).

Questions that are likely to arise as you read the document are handled in the various “Box’s”, so do not distract from the flow of the presentation but answers are there if needed.

Two respondents said that they would give this to staff tasked with developing productivity measures and encourage others to use it. This is an encouraging sign that *Measuring* is well pitched to the intended audience.

The natural tension between simplicity and comprehensiveness has been resolved in favour of comprehensiveness. The case studies help illustrate the recommended process as well as illustrating what might be sacrificed if simplicity is given too much weight. Moreover, the “Takeaways” at the end of each chapter promote easy navigation, allowing the reader to find what they need and (by skipping on) alert them to what they might be missing. On the other hand, for those just interested in applying narrow technical efficiency measures to operational activities, the comprehensive economic approach is unnecessary.

### *3.3 Were the commissioned reports for the inquiry – for example Elizabeth Knopf’s (2017) paper on the History of efficiency measurement by the New Zealand health sector insightful and useful? What about the staff and state sector leader perspectives (Pickens (2017); Plimmer, Cantal & Qumseya, (2017)) – do they add anything of value to the literature on the culture and leadership of the New Zealand state sector?*

For me the most useful insight from the “History” paper is the identification of the two types of measure that have proved enduring:

- efficiency measures that respond to the operational requirements of DHBs (or used strictly for their own learning purposes); and
- health targets that have a high level of legitimacy amongst a number of stakeholders; are regularly monitored and used for performance management; with a well-resourced program to help support the achievement of those targets.

In both cases there are clear motivations for wanting to perform better against the measure because they are seen to be measuring something of value and in both cases there is plenty of effort devoted to helping organisations perform better against these measures.

The staff perspectives are consistent with other measures of employee engagement across the state sector (e.g., as reported in PIF reviews): i.e., individual staff are motivated by their sense of service but do not rate their organisations as highly performing (which suggests a poor goals alignment between staff and organisation and/or that the organisation does not add much value to their staff's efforts). That staff do not rate their organisations as innovative or as developing learning cultures is also consistent with the typical results of the PIF reviews. In short, this data tends to confirm what other data was telling us; so it adds value in that sense.

It is more difficult to say what the leaders perspectives would have added to our understanding of culture and leadership in the NZ state sector, in part because of the wide variation in views.

That said, there were some common themes: that performance was improving but patchy; that the language of productivity was not very motivating; that building Ministerial confidence was important to improving sector performance; that good performance management systems were important to improving performance; that current accountability and performance management requirements imposed from the centre were “costly and fail to provide strong incentives to improve performance”; and, therefore, that a lot of variation in performance came down to variation in the ability of individual leaders.

The lesson that I take from this is that patchy performance that has to rely on good leadership reflects the lack of system-level levers for ensuring that good leadership and strong performance are common and enduring. That may or may not add much to the literature but it does give real support to the focus in the *Improving* report on “System-level changes for a more productive state sector” (emphasis added).

#### *3.4 Does the inquiry promote understanding about the role of innovation as a driver of productivity?*

Yes, although in my view this should have been developed much further (even if relegated to an Appendix in order to preserve the brevity of the report).

The Report rightly notes the limited role for “reallocation” of resources from less to more productive organisations in the public sector and, therefore, that “state sector productivity growth is much more likely to rely on technological diffusion (defined broadly).” (*Measures*, p. 5) Given the central importance of innovation, the subject deserved more attention in at least three respects.

First, a better understanding of the relative importance of generation and diffusion would have been useful. Given the degree of devolution in the delivery of health, education, justice and social services, diffusion is likely to be especially important. In my experience, there is no shortage of innovative practice in these sectors. What stands out is the lack of diffusion of successful innovation

and the timely or widespread adoption of better practice; even for organisations (or even units inside the same organisation, like theatres and wards in hospitals) providing similar services to similar clients. It would be useful to know how common this is across the sectors examined.

Second, given the likely relative importance of diffusion, it would have been useful to contrast cases of fast/widespread adoption and slow/limited adoption: what are the barriers to faster and more widespread adoption and what could be done to accelerate diffusion? There should be enough examples of innovation intended for wide adoption that we should be able to draw some valuable lessons.

There is, for example, a relatively high degree of comfort in the social services towards “piloting” new practice. What has been our experience in encouraging adoption of those that prove successful? There is a widespread view that successful pilots go nowhere and unsuccessful ones are hard to shut down (see, for example, the leaders view reported by Pickens that “pilots were not working as intended” (p.49). Is that true and if so, why and what can be done to address these issues? What distinguishes the exceptions (e.g., what made the health care home concept spread reasonably widely)?

In a similar vein, part of the rationale for private prisons was that public prisons would be able to learn from innovation in the private setting. Has that happened and, if not, why not? Is this a realistic proposition and, if so, what could be done to make more of it?

A better understanding of these issues could be used to support better experimental design and organisational agility; key elements in encouraging both discovery and adoption when the link between services provided and the outcomes desired are not well understood. This unpredictability is common enough in these sectors to deserve serious attention.

Finally, given the Inquiry’s emphasis on measurement, it would have been useful if the Inquiry identified ways of measuring the relative speed and breadth of diffusion of successful innovation.

For example, when is there enough commonality in the client grouping to be able to use large and persistent gaps in performance amongst different service providers (or services provided by a single agency in different locations) as a good indicator that attention need to be given to improvement? Is there a useful enough counterfactual to be able to use the time taken for successful practice (or new drug, device or technology) to become commonplace as a useful indicator? Both these approaches have been used in the private sector to measure diffusion. In terms of the latter, the Report claims that “... measuring diffusion in the state sector is often a relatively straightforward exercise” because administrative data allows direct measurement of changes in practice (*Measures*, p5).

3.5 *Do the barriers to lifting state sector productivity identified in the report resonate with you? How might we have changed or presented them to better resonate with the Treasury and SSC?*

They do resonate: it is hard to argue with the judgements made about the importance or prevalence of each of the seven major barriers to achieving higher productivity identified by the Inquiry. Moreover, it is not immediately obvious that there are any other barriers that would have the same degree of importance or prevalence. In that sense, the identification of these barriers is a significant contribution that the Inquiry has made to our understanding of why state sector productivity seems so poor.

That said, there are two problems with the way they are presented in the *Improving* document:

- As noted in my assessment of “clear delivery of messages” in section 5 below: The barriers are presented in a way that demands attention but does “... little to develop the confidence of the existing workforces in their ability to do much about it.” Indeed, my judgement would be that the way the barriers are presented makes it harder to “... engage the State Sector workforces on the opportunities to do things better” (Terms).
- The focus in the Report is on the importance and prevalence of barriers and not enough on what creates these barriers and so what can be done to address them. That both contributes to the problem in the above bullet point and makes it more difficult to see what Treasury and the SSC might do to address some of the underlying causes.

Take “closed and risk-averse agency cultures inhibit innovation” as an example. While the report does go some way to identifying the cause, it does not go far enough. The report suggests that the adversarial nature of the parliamentary system is at the heart of the issue and that a better balance must be struck between “political accountability on the one hand and creating an environment conducive to innovation on the other.” (*Improving* p. 18). However, it is not clear what would have to change in order to ensure a better balance.

In order to better resonate with Treasury and SSC, this needed to be taken a step further in order to identify what might have to change and what Treasury and SSC could do to help. The discussion in section 3.7 below, for example, points to the need for Chief Executives to earn the confidence of Ministers that their agency can assess and manage the risks involved in innovation. That might better resonate with SSC, for example, because of their interest in the selection and development of Chief Executives and in the development of capacity and capability across the state services.

### *3.6 Would the recommendations, if implemented, improve the ability of state sector agencies to innovate and for innovation to diffuse more rapidly through the state sector?*

The recommendations would, if implemented, improve the ability of agencies to innovate and for innovation to diffuse more rapidly. A small set of very practical and achievable recommendations are designed to gradually strengthen the incentives to be more productive at the same time as helping build measurement capability and capacity in response to that increasing demand.

While the number of recommendations is small, they address many of the barriers identified in the report: e.g., lack of incentives to seek improved productivity and poor measurement capacity and capability.

It is less likely that the recommendations would be sufficient to address all of the identified barriers: e.g., the “hostility” to measurement or the “closed, risk-averse cultures” (see suggestions in 3.7 below for what might be required to help address this latter barrier). However, while probably not *sufficient*, the heavy emphasis on strengthening administrative and budgetary incentives to improve productivity: seems *necessary* to address the identified barriers; is a good place to start; and would stimulate improvement.

The Inquiry teams approach is pragmatic and realistic and could complement other changes to the State Sector currently under consideration, if they are also implemented.

3.7 *Were you convinced by the argument that the measurement of productivity would improve (and the capability to measure would improve) if ministers demanded productivity information? Do you agree that ministers have a role in supporting innovation and making “space” for it?*

While the answer is “yes” the reason is important because, at first blush, this should be the province of Chief Executives.

The State Sector Act makes Chief Executives legally responsible for “the efficient, effective and economic management of the activities of the Department” (State Sector Act s32 (d)) and gave them the freedom to manage.

While Ministers should be able to rely on their Chief Executives to identify and realise productivity gains, they will have a view on how these productivity gains are made. They are likely to object to changes that they see as damaging, too risky or too controversial.

Private sector Chief Executives must also consider the potential for their actions to generate negative publicity and must have an eye to the potential regulatory implications of action that is unpopular. In this regard, the situation of Chief Executives in sectors that are unregulated, regulated or publicly-owned do not vary *in kind* from those in the state-sector. However, there is a steeply increasing difference in *the degree* of sensitivity that Chief Executives must have to public reactions as they move along this spectrum.

That means state sector Chief Executives have to work harder than their counterparts in other sectors to earn the right to innovate. It does not mean that they are unable to innovate and seek productivity improvements. Ministers need to be convinced that state sector Chief Executives can accurately assess and manage the risks associated with innovation, experimentation and change. Gaining the confidence of Ministers that this can be done, and so earning the right to innovate, is all part of the job.

The problem arises when Ministers do not have enough confidence in management or when Ministerial preferences are overly vague or unpredictable. Then Chief Executives start to self-censure in a way that is overly cautious: risk is avoided rather than managed and the risks of doing nothing largely ignored. This leads to the sorts of observations from a variety of groups about the very closed and risk averse nature of the state sector (pages 16-18 of the *Improving* report).

If Ministers were able to be more explicit about the desirability of productivity measures and their level of confidence in their Chief Executive’s ability to assess and manage the risks inherently associated with innovation, experimentation and change then Chief Executives would be less likely to self-censure in an overly restrictive way.

However this does imply a skilled Chief Executive who is able to build that confidence and a Minister that is willing to accept a well calculated and managed degree of risk. The findings in the report about the very closed and risk averse nature of the state sector suggests that this combination is relatively uncommon in practice. That points to the incentives and opportunities facing Ministers as well as Chief Executives. Therefore it would have been useful for the Inquiry to have looked at the incentives facing Ministers, as well as Chief Executives, and what might be done to better align those with the objective of encouraging productive innovation.

*3.8 What's missing from our analysis? (Note: two studies that attempt to get a handle on innovation in the health care sector are still in progress).*

The areas where more could have been done to enhance the value of the analysis have been identified in the relevant sections of this assessment. For example, section 1 identifies the benefits of identifying when measurement is likely to be most helpful and create fewest risks and of suggesting ways of dealing with problems that are typically cited as limitations so as to expand the scope for the application of imperfect "quality adjusted" measures. Section 3.4 above suggests three areas where more attention could have been given to the understanding and measurement of innovation and, especially, the diffusion of innovation.

I have tried to keep these to a minimum, recognising the importance of keeping the document brief and accessible to a wide variety of audiences.

The only other area where more analysis would have been really useful is an assessment of the experience with "raising the bar on new expenditure" (recommendations R 3.6 to R 3.9) and with "improving commissioning (and funding) processes" (recommendation R 3.10).

This set of recommendations are necessary to create the financial incentives that will drive demand for improved productivity and are, therefore, critical. Moreover, the Treasury has accumulated experience with the suggested changes to budget processes and has identified what worked well and what did not. Similarly, some agencies, like NZTA, have very sophisticated commissioning and procurement practices that hold lessons for those looking to shift towards "more results- and outcomes-based systems." (R3.10) There are also useful examples in the wider social and justice sectors of agencies moving in this direction.

For example, the Treasury did experience some issues around the application of "Track 1" in Budget 2017 and suggested that better mechanisms were needed to deal with these issues as well as to better manage bids from NGOs. The enthusiasm with which Treasury will adopt the Commission's recommendations will be tempered by their concerns around these issues. It would have been easier for Treasury to support these recommendations if these concerns had been addressed in the Report.

**4. Effective engagement – how well the Commission engaged with interested parties**

Section 2 commented on the quality of the Commission's process management and the extent to which the scope, balance and type of engagement met the requirements of the Terms and the Commission's own standards that its Inquiries are well-informed and relevant. This section focusses on the quality of those engagements: i.e., the extent to which the Commission was seen to be open, responsive and respectful and, therefore, likely to elicit the best responses from their engagements.

Given the well-deserved reputation of the Commission for the quality of its engagement in past Inquiries, I was surprised by the very wide assessments of interested parties in the quality of engagement from the Inquiry team. Some described the engagement most positively, "Engagement was really good and respectful, they went out of their way to engage both formally and informally and I felt like they were listening and that I was being heard." Others described it in the opposite terms and some were not as interested or engaged as might have been expected.

Some of this variation reflects different degrees of unease in where the Inquiry might lead and the potential for it to add value to what they were doing or to be used against them; so "where people stood" on their willingness to engage positively depended to some degree on "where they sat" in the state sector.

As noted in the section above on process management, the “barriers to lifting state sector productivity” listed in the final report would have also limited agencies enthusiasm to participate actively in the process. It is not surprising then that when their people were engaged by the Commission, their level of engagement was not all that high.

However, these explanations are not sufficient on their own to explain the variation in the quality of engagement with the Commission.

Some of the variation could be attributed to the focus of the Inquiry, which meant that some did not see engagement as a priority for them. Some agencies who risked very little were not as engaged as might have been expected because the narrow focus on “technical efficiency” limited the gains they expected from an application to the sectors in question.

In Health, the DHBs seemed to be much more actively and positively engaged than the Ministry. That may be because the DHBs have operational responsibility for hospitals, are comfortable with narrow productivity measures and enjoy the obvious benefits from increasing hospital throughput.

Education seemed relatively disengaged and this might be explained, at least in part, by the difficulties associated with applying the narrow measure to schools (see discussion in the “Right Focus” section above).

While none of this reflects on the Commission’s approach, there were two instances that hold some lessons for the way the inquiry team engaged with interested parties.

First, both the Treasury and the State Services Commission wanted specific attention given to the benefits, costs and risks of applying the narrow measure of productivity in the sectors concerned; identify where this application was likely to be most beneficial and most problematic; and what might be done to reduce the costs and the risks. The Final Report does address these issues to some degree, in particular the preconditions for using productivity measures for performance management. However, more could have been done to address these issues and doing so may well have increased the scope for the useful application of quality adjusted service measures.

Second, some of the most negative criticism came from people enthusiastic about productivity measurement and improvement and actively engaged in furthering both. Some of this seemed to reflect a concern that they might lose control around the presentation of what they were doing. While understandable given the importance of issues like context, measurement, motivation and potential use, this does suggest that they did not rank the potential additional benefits of engagement with the Commission all that highly and saw some risk in the way the Commission might present their work. At the very least, the Inquiry team was unable to allay those concerns.

Interestingly, some of their criticism also seemed to reflect a difference between an operations approach and an economic approach. Given that a narrow technical efficiency focus is likely to be most usefully applied to operational activities, it might have been helpful to have an operations person on the inquiry team and to frame more of the conversation in operational (versus economic) terms. This may also have encouraged more agencies to share more of their experiences in using technical efficiency/throughput measures to improve their productivity. When I asked one of the inquiry team members what had surprised them most about the whole exercise they said “the number of agencies who were doing a lot more than they were getting credit for but who were reluctant to share it with us.”

## 5. Clear delivery of messages – how well the work is communicated and presented in the final report

The Commission wants its work to be “accessible”, so it needs to be presented and communicated in a way that is easy for a variety of different audiences to understand.

In this case the “final report” should be seen as a combination of five separate documents: Final Report: Improving state sector productivity; Final Report; Measuring state sector productivity; Cut to the chase summary, A3 summary and Media Release.

Each of these will appeal to different audiences and so, together, maximise the likely reach of the material. In particular, the *Measuring* document is written for “... individuals and teams within the state sector who are intending to develop productivity measures.” (p iii) This is written as a guide for this audience and is intended to be more enduring than the *Improving* document, which is aimed at senior ministerial advisors and agency managers; especially officials in Treasury and SSC who will need to implement nine of the eleven recommendations. Each document is independent and so can be read and understood without having to read the other.

Both the *Measuring* and *Improving* documents have a form of summary at the end of each short chapter: “takeaways” in the former and “findings” and “recommendations” in the latter. This makes it easy for people to navigate their way through the documents and to find the material most relevant to them. These chapter summaries also offer a more permanent form of summary for those who only need to have a high level view of what the documents are saying.

Communication and presentation is greatly aided by the relative brevity of the documents and the relatively small number of recommendations: the two “Final” documents are about 50-55 pages each and the *Improving* document only has 11 recommendations.

The inquiry team has taken a very pragmatic and realistic approach to developing a small set of mutually reinforcing recommendations and the recommendations themselves are practical and achievable.

Accessibility is further enhanced by the liberal use of boxes, case studies, graphs and tables which break up the text and make the documents more interesting and accessible to those with different interests.

The only thing that may have enhanced the “accessibility” of the work would be to use more of an operational approach and language in describing the measurement and use of the narrow “technical efficiency” measures of productivity highlighted by the Terms (see discussion in section 4 above).

One of the aims of the Terms is to “engage the State Sector workforces on opportunities to do things better.” This is a much tougher test because it requires that the material not only be accessible to different audiences but also motivating to state sector workforces.

The description of the “barriers to higher state sector productivity” describe a state sector that does not make good use of the information available; is closed and risk-averse; has poor policy and commissioning practice; employs restrictive rules and funding models; with patchy monitoring and evaluation; with some people actually hostile to the concept of productivity and resistant to its measurement. Taken together, these barriers describe a situation that demands attention. Unfortunately they also do little to develop the confidence of the existing workforces that they will successfully overcome these barriers. The Report paints a picture of a mountain to climb, with a

substantial prize at the top, but with many on the climbing team that you would not want to take on the journey. That is unlikely to inspire their engagement.

#### **6. Overall quality** - *the overall quality of the inquiry taking into account all factors*

There are a number of factors outside the Commissions control that have complicated this inquiry. The Terms required application of narrow technical efficiency measures to sectors where broader measures were required in order to meet the overall aims set out in the Terms. Few of the directly affected Ministries and departments displayed much interest in engaging proactively with the inquiry team. The barriers that need to be overcome to make significant headway in improving state sector productivity are substantial. The Government that commissioned the inquiry was replaced during the inquiry process. New Governments have lots of other priorities, all demanding officials' time and attention during the latter part of the inquiry process.

While these factors would have weighed on any inquiry, the Commission has produced an important and high quality contribution to our understanding of state sector productivity, its measurement and the barriers facing productivity improvement. It has provided useful guidance to officials who are interested in productivity measurement. Perhaps most significantly, it has identified a few practical, achievable and necessary recommendations that, if implemented, would go a long way to improving state sector productivity.

#### **7. Having Intended Impacts** – *what happens as a result of this work?*

Delivering on the three aims in the Terms requires more attention on improving productivity and its measurement. That, in turn, requires strengthening demand for improved productivity and building the capability and capacity to respond to that demand. These are the “intended impacts”.

The recommendations in the *Improving* report are practical and implementable and, if accepted, would make an important contribution to these objectives. While the *Measuring* report will be voluntarily picked up by a few officials keen on developing better measures, this is unlikely to be either a common or enduring response if the recommendations aimed at strengthening demand for productivity improvement are not adopted by government.

On one hand, it is hard to see why any government would not want to reap the benefits of improved state sector productivity: especially given the sheer size of the sector in the economy and importance in our lives. Taxing more is always controversial, so getting more out of the taxes already being raised must be attractive. It seems clear from the Inquiry that there is considerable scope to do that over time. Moreover, implementing the Commission's recommendations is unlikely to attract much public attention, let alone controversy and would be phased to ensure demand and capacity are reasonably matched. While promises of improved productivity can be met with cynicism, taxpayers may well appreciate the effort.

Clearly, significant improvements will take time, so implementing the recommendations is unlikely yield much immediate benefit. That may make this less of a priority than other claims on Ministers time. However, the upfront investment in time and effort is not all that significant either.

On the other hand, the Government has changed since the Inquiry was commissioned, so may have different priorities to the previous Government, including different priorities for improving state sector performance. Moreover, the Beehive has not welcomed or even acknowledged the Report, which contrasts with its response to the Report on Emissions which was released about two weeks

later. The PSA did respond, saying that they did not dispute that productivity growth in the state sector appears poor but that they were “wary” of the Commission’s proposed remedies.

While it would be foolish to try and predict the current Government’s response, these initial indicators are not encouraging. Without Ministerial support it is unlikely that Ministerial expectations will be changed in the way the Report recommends or that the Minister of Finance will push hard on the three budget changes recommended, unless he wants to do so for other reasons. Without these changes, there is little additional incentive to strengthen demand for either productivity measurement or improvement.

Lack of Ministerial interest and continued weak budgetary incentives will inevitably be reflected at the official level, where any changes will need to be initiated and implemented by officials. Chief Executives are legally responsible for “the efficient, effective and economic management of the activities of the Department” (State Sector Act s32 (d)) and have the freedom to do so. However, these are long standing arrangements and it is hard to see agencies with little inherent interest in the recommendations picking them up voluntarily (see also section 3.7 above). Those that are inherently interested are likely to be taking action anyway, albeit they may well benefit from insights in the *Measuring* report.

Most of the remaining recommendations are directed at either the Treasury and/or the SCC. Some of these recommendations fit well with their roles and may well be picked up; especially those aimed at strengthening capacity and capability across the public sector. However, as noted above, in the absence of the measures aimed at increasing the demand for productivity improvement, it is hard to see any new initiatives from central agencies having the impact that they would otherwise have had.

## Annex A

### Terms of Reference for review of State Sector Productivity inquiry

#### Background

This independent review of the *Measuring and improving state sector productivity inquiry* is a valuable opportunity for the Commission to learn from a seasoned operator about what could have been done better, both with respect to the inquiry process and the final reports. The inquiry reports are intended to stand on their own, and you should evaluate them with that in mind. You can view other external evaluations from our other inquiries on our website.

#### Deliverables

The deliverable is a report of your review of the Commission's *Measuring and improving state sector productivity inquiry*. The deliverable would be a relatively short report summarising the evaluation that we could publish on our website, quote in reporting our performance (e.g. on our website and for use in the Annual Report) and use in improving our performance.

Our inquiries are formally evaluated on the following performance measures:

- **Right focus** – the relevance and materiality of the final inquiry reports in meeting the TOR;
- **Good process management** – the timeliness and quality of the inquiry process;
- **High quality work** – the quality of the analysis, guidance and recommendations in the final reports;
- **Effective engagement** – how well the Commission engaged with interested parties;
- **Clear delivery of messages** – how well the work is communicated and presented in the final reports;
- **Overall quality** – the overall quality of the inquiry taking into account all factors.

While these criteria are all important (and required as part of our formal reporting) the Commission is particularly keen to get your assessment of the quality of the analysis leading to the recommendations and the quality of the guidance on measurement. We would also appreciate your views about the “value add” from the Commission's work on this inquiry. For example:

- Has the inquiry, in your view, advanced current thinking on the measurement of state sector productivity at a service level (Dunleavy (2015), Dunleavy & Carrera (2013), Van Dooren (2005)) by placing measurement within an institutional and performance management context?
- Did we pitch *Measuring state sector productivity* at the right level for its intended audience? Is it simple and comprehensive enough to guide productivity measurement in the state sector?
- Were the commissioned reports for the inquiry – for example Elizabeth Knopf's (2017) paper on the history of efficiency measurement by the New Zealand health sector insightful and useful? What about the staff and state sector leader perspectives (Pickens (2017); Plimmer, Cantal & Qumseya, (2017)) – do they add value to the literature on the culture and leadership of the New Zealand state sector?
- Does the inquiry report *Improving state sector productivity* promote understanding about the role of innovation as a driver of productivity?

- Do the barriers to lifting state sector productivity identified in the report resonate with you? How might we have changed or presented them to better resonate with the Treasury and SSC? Would the recommendations, if implemented, improve the ability of state sector agencies to innovate and for innovation to diffuse more rapidly through the state sector?
- Were you convinced by the argument that the measurement of productivity would improve (and the capability to measure would improve) if ministers demanded productivity information? Do you agree that ministers have a role in supporting innovation and making “space” for it?
- What’s missing from our analysis? (Note: two studies that attempt to get a handle on innovation and its diffusion in the health care sector are still in progress).
- 

Note that the Commission’s performance framework also contains another dimension:

- **Having intended impacts** – what happens as the result of the Commission’s work

While it is likely too early to judge this aspect, you should make any observations that you feel you can make.

The review should note any lessons that can be taken and make recommendations for any future improvements to the Commission’s inquiry processes and reports.

The report must also contain a ‘summary assessment’ (or alternate name) that summarises your perspective on each of the performance dimensions (a short paragraph on each) – this is useful for the Commission’s Annual Report.

The approach in undertaking the evaluation is up to you but we can assist by making available all relevant papers/documentation and communications material. We can also assist in identifying and coordinating discussions/interviews with key people (Commissioners, NZPC staff and external stakeholders) if you wish.

## **Annex B**

### **Minister of Finance**

#### **Terms of Reference - New Zealand Productivity Commission Inquiry into State Sector Productivity**

Issued by the Minister of Finance (the “referring Minister”). Pursuant to sections 9 and 11 of the New Zealand Productivity Commission Act 2010, I hereby request that the New Zealand Productivity Commission (“the Commission”) undertake an inquiry into how the New Zealand State sector can effectively measure and improve productivity in core public services, with a particular focus on health, education, justice, and social support.

#### **Context**

Improving the productivity of the State sector, the value we are realising from our resources, helps improve the prosperity of the country, and allows for better outcomes to be achieved from scarce tax payer resource.

Recent progress has been made in improving value across the different dimensions of value for money performance. The Better Public Services Results determine priority areas for improvement. Social Investment and other effectiveness work is getting better at identifying where to invest and tracking what the impact of investment is.

A third dimension of performance is efficiency/productivity. For many of the core public services that constitute a large proportion of existing expenditure, there are still opportunities to better understand efficiency and how to optimise inputs/resources in delivering quality products and services. Current gaps in good measures of productivity limit assurance Ministers have on performance and innovation of current delivery models, and Chief Executives ability to understand and improve their business. It also suggests an opportunity to achieve more from current resources, and better engage the State sector workforces on opportunities to do things better.

Public services are often complex covering a range of services, clients, and different mechanisms to achieve a range of desired outcomes. This can make it more difficult than private sector industries to capture performance, and to take actions to improve it. Internationally, there are few common productivity measures that capture quality dimensions in key sectors like education and health. But, there are lessons on how to better understand dimensions like quality in inputs and outputs, leverage innovation and economies of scale, and improve productivity and efficiency in the public sector.

#### **Scope**

The Productivity Commission (the Commission) is to consider New Zealand and international public and private sector best practice in understanding and improving productivity. This should focus on the narrower definition of productivity as how efficiently inputs/resources are being utilised to generate quality outputs/services.

The Commission should take account of broader definitions of performance and productivity, in considering how to capture elements like quality, and how efficiency measures can complement dimensions like effectiveness. However, the Commission should not focus advice on the contribution of services to longer-term outcomes, prioritisation of interventions, or other performance dimensions already being developed through social investment or other work programmes.

The inquiry should focus on developing practical guidance and recommendations that consider perspectives and roles of different state sector decision-makers such as Ministers, Chief Executives, and managers, and how these different needs can be balanced.

The inquiry is to focus on guidance that is relevant to decision-makers across the “core” services in the health, education, justice and social development sectors, such as: teaching, hospitals and primary healthcare, policing, courts, corrections, and work and income services.

Having regard to the above, the Commission should undertake an inquiry that considers and provides advice on:

- a) How to measure efficiency/productivity in each of the identified core public service sectors: health, education, justice, social support. This should focus on meso (sector) and micro (function or service) level measures. Guidance should consider key measurement and accuracy issues, and how imperfect measures are most appropriately and usefully employed.
- b) The appropriate role of identified efficiency/productivity measures in public sector performance frameworks, with the goal of improving assurance to Ministers and incentives on agencies for improvement. This should draw on theory and evidence of incentive and disincentive effects of measurement and other performance approaches on different workforces.
- c) Developing the capability, culture and systems that can support agencies to better measure, understand and improve productivity.

The Commission should prioritise its effort by using its judgement as to the degree of depth and sophistication of analysis it applies to satisfy each part of the Terms of Reference; and to the degree of depth in each specific sector, while providing advice on best measures in the identified sectors.

### **Exclusions**

The Commission should not carry out in depth analysis or provide detailed recommendations on specific policies relating to service access or provision in sectors.

The Commission should not duplicate work on issues like where to invest, or service effectiveness, being developed as part of the social investment approach.

### **Consultation requirements**

In undertaking this inquiry the Commission should consult with key interest groups and affected parties relevant to the identified sectors and particular services where efficiency measures are identified. Consultation should include public sector agencies, those in receipt of public services, and private sector agents who may have relevant insights.

### **Timeframe**

The Commission must publish a draft report and/or discussion document, for public comment, followed by a final report that must be presented to the Minister of Finance as Referring Minister by 30 August 2018.