Proposed framework for reporting on ‘what works’ to improve service outcomes

The Steering Committee for the Review of Government Service Provision is seeking feedback on the proposed framework for reporting on what works to improve service outcomes (this paper).

The proposed reporting framework is intended to provide a practical approach to identify from existing research what works to improve service outcomes in the areas covered by the Steering Committee’s performance monitoring report, the Report on Government Services (RoGS). For more information on the RoGS see:

http://www.pc.gov.au/research/ongoing/report-on-government-services

Your feedback

Feedback on this paper can be sent by email or post to the Secretariat for the Steering Committee for the Review of Government Service Provision (based in the Productivity Commission). The opportunity to provide feedback on this paper is open from Thursday 19 May to **Thursday 14 July 2016**.

By email: gsp@pc.gov.au

By post:
Draft reporting framework
Government Performance Reporting and Analysis Branch
Productivity Commission
Locked Bag 2, Collins Street East
Melbourne VIC 8003

Q and A

If you would like any further information about the proposed reporting framework or the RoGS, or have questions about this process, please contact the Secretariat at: gsp@pc.gov.au

Summary

Background

Performance reporting (such as the Report on Government Services [RoGS]) provides information on the performance of jurisdictions in delivering outputs and outcomes, but it does not reveal the impact of different inputs or how they should be used to influence outcomes. Without analysis of causal links, long-term and cost-effective improvements in outcomes are only likely to be achieved through trial and error. A more rigorous and nuanced evidence base would enable governments to draw on those policies/programs/interventions that work, and learn from those that have not worked, and so deliver services more efficiently and effectively.[1] This is critical, with government recurrent expenditure on the services reported in the RoGS around $190 billion annually.

Proposal

The proposed reporting framework on what works to improve service outcomes, is based on an evidence 'pipeline' (supply chain approach) with a feedback loop (see figure 1).[2]

The 'pipeline' moves sequentially from:

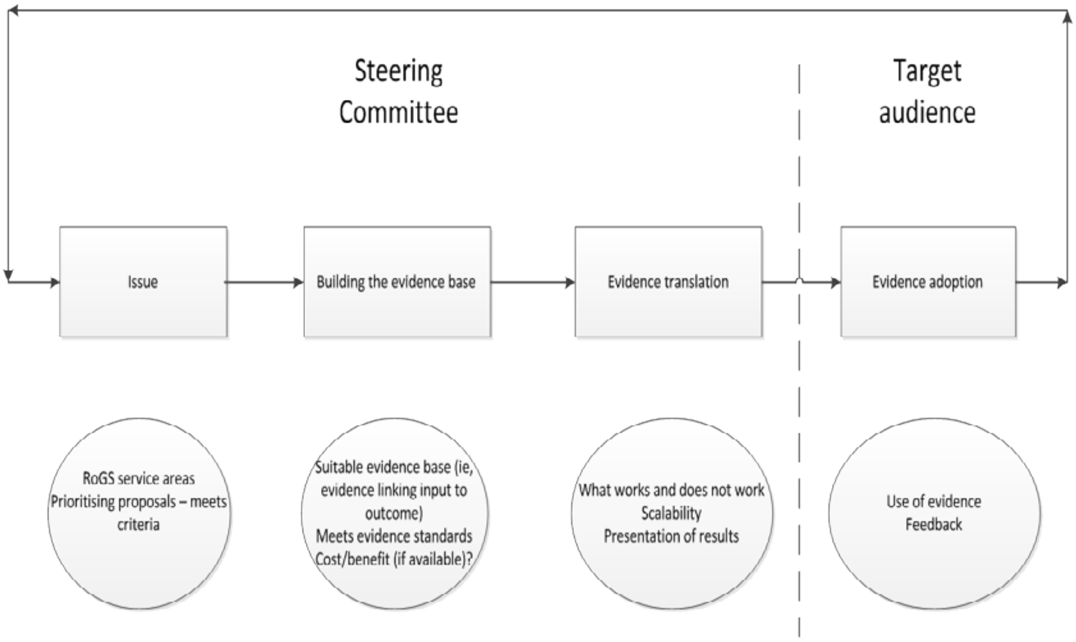
- the Steering Committee determining the issue for analysis, with the threshold being robust evidence directly linking changes to the input/output being assessed to changes in the service outcome, and prioritisation of areas of pressing social need, major public spending, minimal improvement over time to the relevant RoGS indicator/area and an existing research base (areas with limited or no research linking changes to an outcome will be reported as an evidence gap)
- the Secretariat synthesising existing evidence (national and international), where the evidence meets agreed standards, including reporting on evidence gaps (ie, reporting where there is a gap following the synthesis — this could be due a lack of evidence, poor quality of evidence and/or gaps when compared to evidence standard)
- the Secretariat reporting on what works (*and* what does not work) to facilitate improved service outcomes, including scalability of the interventions, with presentation of results tailored to the target audience (which includes policy makers, governments, service providers, researchers and clients)
- the Secretariat following up with the target audience on adoption/use of the evidence, which in turn feeds into subsequent reporting.

---

[1] In the United States, the Washington State Institute for Public Policy provides investment advice to legislators on what is proven to work. It has been estimated that it saves the state over $1b every two years (cited in HM Government 2012).

[2] The 'pipeline' model was originally developed by Glasziou and Haynes (2005) in the healthcare field to outline the different stages that define research use from a practitioner perspective.

Figure 1 **Proposed reporting framework on what works to improve service outcomes**



Two key aspects of the proposed framework are:

- the use of existing research
  - There is a substantial research base that exists but in many cases is untapped. Drawing on this research and consulting with leading researchers will enable identification of what works and key research gaps, to target where individual evaluations may be required.
- reporting on adoption/use of the evidence on what works
  - If the ‘what works’ reporting is not used by the target audience it will not achieve its aim to facilitate improved service delivery. The Secretariat will investigate *if* the outcomes from this reporting have been used (nationally or internationally), but will not make an assessment in relation to *how* they were used.[3]

The proposed reporting framework is at **attachment 1** and includes a worked example for illustrative purposes.[4]

---

[3] Assessment of implementation is outside the scope of the Steering Committee’s role.

[4] The worked example is not a definitive proposal – it is illustrative only, to demonstrate application of the proposed reporting framework.

Attachment 1: Framework for reporting on what works to improve service outcomes

Objective of the reporting

The objective of the reporting framework is to produce accessible information on what works to improve outcomes in RoGS service areas[5] that will be used by the target audience (governments, policy makers, service providers, researchers and clients).

Target audience for reporting

Many stakeholders are involved in services covered in the RoGS including: government, policy makers, regulators, service providers, researchers and clients. The Secretariat suggests the target audience for 'what works' reporting include:

- policy makers — in coming to an informed view of what is and is not cost-effective in these services
- governments — to inform decisions on how best to spend public money
- service providers — in establishing how best to deliver services and how to improve services (includes government and non-government providers)
- researchers — so they are more acutely aware of the nexus between research and policy outcomes
- clients — who increasingly have agency in delivering and crafting interventions, as in consumer-directed care.

Framework

The framework should enable reporting in a way that is accessible to the target audience. This should include reporting on tangible levers that could be used by the target audience to make a measurable change in the relevant outcome area.

The proposed framework has a sequential order with each element following the one preceding (refer to figure 1 on page 3). The elements of the reporting framework, outlined in more detail below, are:

1. Determine issue for assessment
    (a) selecting the service aspect for assessment (what is the area being tested), with a threshold of demonstrated link of input/output to service outcome (if no evidence, reporting the evidence gap)

---

[5] RoGS covers: childcare, education and training; justice; emergency management; health; community services; housing and homelessness.

   (b) prioritising assessment proposals based on areas sharing the common characteristics of pressing social need, major public spending, minimal improvement over time in relevant RoGS indicator/area and an existing evidence base

2. Building the evidence base, drawing on existing evidence sources
   (a) selecting evidence sources, drawing on national and international research
   (b) application of agreed evidence standard (assessing risk of bias, relevance of study aspects and risk of harm from intervention)
   (c) costs and benefits (including reporting on evidence gaps)
3. Evidence translation, to reach target audience
   (a) what works and what doesn't work
   (b) scalability (to national level or significant level to have national impact) and persistence
   (c) costs and benefits
   (d) presentation of results for different audiences
4. Evidence adoption[6]
   (a) use of the evidence
   (b) feedback on what would improve usability

1. Determine issue for assessment

(a) Selecting the service aspect for assessment

Using the program logic model adopted in the RoGS, the focus should be on inputs/outputs that have a demonstrated link to the service outcomes (so that a change to the input/output will imply a subsequent change to the outcome).

This link will draw on existing research (national and international) and may be determined by evaluation (direct link) or synthesis of a sufficient volume of high quality research studies (indirect link). The synthesis may be pre-existing, or undertaken as part of this reporting exercise (resources permitting). Synthesis will be reviewed against the following characteristics: risk of bias; consistency; precision; directness; reporting bias (IOM 2011).

Where a service area does not have a demonstrated link for the intervention to the desired outcome, this should be reported as a research gap.

The Secretariat will seek feedback from the target audience on potential topics for consideration and short list to those linked to RoGS as per specifications above.

---

[6] Whilst evidence adoption is outside the remit of the reporting process, it is an important part of the feedback loop on the accessibility and/or usefulness of what works reporting. The objective of improved service delivery can only occur if the evidence is acted upon.

[A worked example is provided to illustrate how each phase of the framework is incorporated. The example is for illustrative purposes only and does not represent the Secretariat's view of interventions that should be assessed nor does it contain the depth of assessment required for a published review.]

Example (part 1):

Pre-existing synthesis of research indicates cancer screening (output of health service) enables earlier detection and treatment which can reduce cancer death rates (long term outcome of health services) (Pierson et al 2013). Nationally in 2012, there were 869 new cases of cervical cancer diagnosed. In 2013, 224 women died from cervical cancer. Rates have been similar for the last decade.

(b) Prioritising proposals

Based on the work completed in the UK for the national What Works centres (Alexander and Letwin 2013), the following criteria should be used in prioritising multiple proposals:

- sectors of pressing social need
- major public spending
- minimal improvement over time in relevant RoGS indicator/area
- where there is an evidence base, but limited synthesis and/or communication of the evidence.

The Secretariat will short list and prioritise proposals for Steering Committee decision.

Example (part 2):

In Victoria in 2013, of the 139 women diagnosed with cervical cancer, 49 (40 per cent) were lapsed screeners (last screened more than 2.5 years) (VCCR 2014). Nationally in 2012-13, only 3 in 10 women responded within 3 months to their 2-yearly reminder letter (AIHW 2015). What works to improve screening rates?

2. Building the evidence base

(a) Selecting evidence sources

The first step in this process is the selection of appropriate evidence sources. What constitutes 'good evidence' is contentious. Nutley, Powell and Davies (2013) provide a matrix comparing the appropriateness of different research designs to different research questions, which illustrates that there is no one approach that works for all. In many areas, randomised control trials (RCTs) provide the 'gold standard' for assessing the impact of different interventions (allocation to different groups, similar on both observable and non-observable characteristics[7]). However, there are currently few RCTs in Australia, as they can be costly and difficult to implement, may involve significant ethical dilemmas, and do not necessarily address the multiple dimensions that may affect the outcomes of interventions.[8] There are alternative quasi-experimental methods using quantitative techniques and natural experiments that can provide major insights. Further, questions about how the intervention works and whether it is right for different groups are better answered using qualitative research (Nutley, Powell and Davies 2013).

A matrix approach will be adopted for the selection of evidence sources, drawing on national and international research, and applying the agreed Evidence Standard (see below).

(b) Application of agreed Evidence Standard

Drawing from the US What Works Clearinghouse Procedures and Standards Handbook (US DoE IES WWC 2013), for each individual study the following elements would constitute the evidence standard:

- assess the risk of bias (including sample attrition and confounding factors[9])
- assess relevance of the study's populations (including sufficiently large and representative sample), interventions, and outcome measures (outcomes should show evidence of face validity and reliability)[10]
- use observational studies to evaluate potential negative side-effects of interventions[11].

The weight applied to each element above may differ where multiple studies, rather than a single study, are being assessed (for example, relatively smaller samples may be sufficient where there are a large enough set of studies showing a similar direction in results).

Processes that support the quality of the results, such as the transparency of methods and diagnostics, are also important. Where appropriate, expert refereeing may also be used.

---

[7] Examples of observable characteristics include age and sex. Examples of non-observable characteristics include motivation and capability. Both matter when they influence the intervention selection and outcomes.

[8] See page xvii in BMJ (2004) for cases where RCTs are unlikely to answer the research question.

[9] Sample attrition here relates to bias in the loss of sample members/groups. Confounding factors refer to a factor not accounted for that is correlated with both the intervention and the outcome and is present for all members of one group but not another group, meaning that you can't tell if the intervention or the confounding factor influenced the outcome.

[10] Face validity is demonstrated for an outcome measure where it is clearly defined and directly measures the construct it was designed to measure. Reliability is the degree to which an assessment tool produces stable and consistent results.

[11] Observational (cohort and case control studies) studies are used to identify if there are any unintended consequences that could occur which need to be taken into account. RCTs may provide inadequate data on harms due to short follow-up periods or have limited applicability to real-world settings (US DoE IES WWC 2013). This blend of quantitative and qualitative results is an example of 'triangulation' of research.

(c) Costs and benefits

Assessment of costs and benefits are important for governments to make informed decisions on expenditure, particularly where there are competing requests for funding.

Very few evaluations incorporate a cost-benefit analysis. Cost-benefit analysis will be reported where available, and noted where not available. Resources permitting, where it is possible to reach a judgment on costs and benefits, this will be reported.

Example (part 3):

NSW Department of Premier and Cabinet (DPC) Behavioural Insights Unit (BIU) ran a RCT to test which of five reminder letters (including the current letter as control) were more effective at prompting women to have a Pap test (75,000 letters sent to women in NSW). The most effective letter performed statistically significantly better than the control (32.2 per cent compared to 29.7 per cent) – which could lead to an additional 7500 women having Pap tests within three months of their reminder letter. The trial has only been conducted once – NSW DPC BIU is intending to replicate the trial with subgroups of women. A formal cost-benefit analysis has yet to be completed. Given that the costs of sending a randomised group of letters is very low, prima facie the trial would pass a cost-benefit test to the extent that Pap tests were themselves efficacious.

3. Evidence translation

(a) What works and what doesn't work

'What works' means the intervention resulted in a direct improvement in the output or outcome as per the evidence standard in section 2 above. Historically, reporting has focused on the statistical significance of the results. The Secretariat recommends giving most weight to the magnitude (effect size) of the results[12] and reporting the confidence intervals for these results, and their policy relevance. [The combination of effect size and statistical significance is currently used by ACARA to estimate the nature of the difference in NAPLAN results for student learning outcomes.]

For results that do meet the evidence standard in section 2, but do not show significant magnitude nor statistical significance in improvement in outcomes following intervention, these should be reported as what does <u>not</u> work[13]. Reporting what doesn't work is a critical component to evidence-based reporting to address issues of direct replication of policies and programs that have been shown through robust assessment to not work.

---

[12] Ziliak and McCloskey (2012) provides a detailed read on the need to consider 'actual' as well as 'statistical' significance (ie, size does matter, not just significance). By reporting what works covering actual and statistical significance we can better present the costs and benefits of the intervention.

[13] In reporting what does not work it is important to note the specifics of the evidence source used, to enable changes to be made and further testing conducted. For example, an intervention may be later modified and then found to work.

In communicating with the target audience on what works, it will be important to also communicate the nature and level of risk from implementation failure.

(b) Scalability

In most instances, experimental studies such as RCTs are conducted on relatively small samples and scalability should also be considered before considering national (or larger scale) applicability. A common criticism of government is a successful small-scale pilot being rolled out nationally without the same success.[14]

The Secretariat recommends a mix of three approaches to assess the scalability of an intervention:

- *Replication*: where the intervention has been replicated in different settings for different samples.
- *Assessment*: where the intervention has not been replicated, its scalability should be assessed against the following criteria (drawn from the NSW Ministry of Health guide on scalability (Milat, Newson and King, 2014)):
  - effectiveness (achievement of significant intervention effects)
  - potential reach and adoption (extensive enough to have a population impact)
  - alignment with strategic context (intervention is consistent with national, state and/or regional policy directions)
  - acceptability and feasibility (could the intervention be realistically scaled up given known information on costs, workforce and infrastructure requirements and acceptability to stakeholders).

  This will require nuanced understanding of the context of the intervention and may require engagement with key researchers.
- *Survey*: where the intervention has been conducted within one jurisdiction, other jurisdictions will be surveyed (informed by the assessment undertaken by the Secretariat at the ‘Assessment’ step above) to seek their feedback on applicability to their jurisdiction[15].

If the intervention is determined to *not* have a sufficient level of scalability it will not be included in ‘what works’ reporting, but reported to note this limitation.

---

[14] For example, Head Start, a federal early childcare program in the US. A 2001 review found positive long term effects, but acknowledged the lack of a high quality RCT. In 2010, the results from a large high quality RCT were less encouraging showing impact diminished over time and negative outcomes for some cohorts (US DoHHS, 2010).

[15] This survey approach was previously used by the Steering Committee in the mid-90s to assess scalability of service delivery reforms.

(c) Presentation of results

There is extensive literature on the gap in knowledge translation from research to policy/practice. For results to be used they need to be presented in short, accessible formats with minimal statistical jargon, and with multiple dissemination channels to promote findings including social media (Shepherd 2014).

To maximise the translation of evidence into practice, research by Grimshaw et al (2012) found that translations needed to come from a credible source, identify the key messages for the different target audiences, and fashion the language and products accordingly.

The specific presentation of results will be informed by consultation with the target audience, and could include a mix of report summaries, blogs by trusted figures in relevant policy areas, research portal and multimedia site (Makkar et al 2015).

Example (part 4):

The most effective letter performed statistically significantly better than the control (32.2 per cent compared to 29.7 per cent) – which could lead to an additional 7500 women in NSW having Pap tests within three months of their reminder letter.

[Scalability was not assessed – this will be of particular interest to jurisdictions with different population groups, such as the NT with a relatively large Indigenous population and younger age profile]

The NSW DPC BIU released a short infographic on the intervention which was very easy to understand and referenced in BIU blogs. It is not known if the results were targeted to particular groups.

4. Evidence adoption

(a) Use of evidence

Assessment of use will be undertaken by:

- analysing views to reporting pages on the PC website using Google analytics
- researching references to reporting on external websites
- seeking feedback from the target audience on adoption/use and any barriers (see below)
- regular reporting to governments on implementation.

(b) Feedback

Tseng (2012) states the importance of engaging with the target audience to gain a deeper understanding of how they acquire and use the research, which can enhance relevance and accessibility of future research.

The public consultation is the first phase in engagement with the target audience to seek their feedback on the proposed reporting framework. A regular follow up with the target audience will be required following the release of new reports.

Distinguishing features of this reporting framework

For this reporting to be useful it should add value to existing national reporting exercises. The proposed framework has the following distinguishing features:

- scope will include all RoGS service areas (existing reporting is specific to either a topic, sector or demographic group)
- will focus on a specific aspect of a service that has a causal link to the service outcomes (existing reporting may focus on aspects that do not have evidence of a causal link)
- will draw on national and international evidence that meets an agreed evidence standard (existing reporting may focus on one evidence source, or use evidence of variable quality)
- will consider tangible levers for governments/policy makers/providers to directly influence service outputs and outcomes (existing reporting may focus on aspects of services that should be influenced, but not how they can be influenced)
- if not national, will have scalability to have significant impact at the national level (existing reporting is often focused on program/policy, but assessment for scalability is not often considered)
- include reporting on what works and to what extent, and reveal what doesn't work (the latter is important to prevent 're-inventing the wheel' and improving efficiency in service provision, but is rarely reported)
- where possible, will include costs and benefits for what works
- seeks to engage with the target audience to ensure evidence translates into policy and practice.

References

AIHW 2015, *Cervical screening in Australia 2012-13*, http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=60129550872 (accessed 17 March 2016).

Alexander, D. and Letwin, O. 2013, *What Works: evidence centres for social policy*, http://dera.ioe.ac.uk/17396/1/What_Works_publication.pdf (accessed 26 March 2016).

BMJ 2004, *BMJ Clinical Evidence, Issue 11, June 2004*, BMJ: London.

Glasziou, P. and Haynes, B. 2005, The paths from research to improved health outcomes, *Evidence Based Nursing* 8, no. 2, pp. 36-38. http://ebn.bmj.com/ content/8/2/36.full (accessed 1 April 2016).

Grimshaw, J. M., Eccles, M. P., Lavis, J. N., Hill, S. J. and Squires, J. E. 2012, Knowledge translation of research findings, *Implementation Science* 2012: 7:50

http://implementationscience.biomedcentral.com/articles/10.1186/1748-5908-7-50 (accessed 18 March 2016).

HM Government 2012, *The Civil Service Reform Plan, June 2012*, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/305148/Civil-Service-Reform-Plan-final.pdf (accessed 26 March 2016).

IOM (Institute of Medicine) 2011, *Finding What Works in Health Care: Standards for Systematic Reviews*, Washington, DC: The National Academies Press.

Makkar, S. R., Gilham, F., Williamson, A., and Bisset, K. 2015, Usage of an online tool to help policymakers better engage with research: Web CIPHER. *Implementation Science : IS*, *10*, *56.* http://www.ncbi.nlm.nih.gov/pmc/ articles/PMC4419443/ (accessed 26 March 2016).

Milat, A. J., Newson, R. and King, L. 2014, *Increasing the scale of population health interventions: a guide*. NSW Ministry of Health: Sydney. http://www.health.nsw.gov.au/research/Publications/scalability-guide.pdf (accessed 18 March 2016).

Nutley, S.M., Powell, A.E. and Davies, H.T.O. 2013, *What counts as good evidence*, Alliance for Useful Evidence, London. http://www.alliance4usefulevidence.org/ assets/What-Counts-as-Good-Evidence-WEB.pdf (accessed 26 March 2016).

Peirson, L., Fitzpatrick-Lewis, D., Ciliska, D. and Warren, R. 2013, *Screening for cervical cancer: a systematic review and meta-analysis*, System Reviews 2013 2:35.

Shepherd, J. P. 2014, *How to achieve more effective services: the evidence ecosystem*, What Works Network/Cardiff University, Cardiff.

Tseng, V. 2012, The uses of research in policy and practice, *Social Policy Reports*, 26 (2), pp. 3–16.

US DoE IES WWC (U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse) 2013, *What Works Clearinghouse: Procedures and Standards Handbook (Version 3.0)*, http://ies.ed.gov/ncee/wwc/pdf/ reference_resources/wwc_procedures_v3_0_standards_handbook.pdf (accessed 18 March 2015).

US DoHHS (U.S. Department of Health and Human Services, Administration for Children and Families) 2010, *Head Start Impact Study: Final report*. Washington D.C. http://www.acf.hhs.gov/sites/default/files/opre/executive_summary_final.pdf (accessed 29 March 2016).

VCCR (Victorian Cervical Cytology Registry) 2014, *Statistical Report 2014*, http://www.vccr.org/site/VCCR/filesystem/documents/dataandresearch/StatisticalReports/VCCR_StatisticsReport_2014_Digital_FinalApproved.pdf (accessed 18 March 2016).