# 10.0
# Evaluation

*The terms of reference for this inquiry asked us to investigate and report on a framework for, and best practice approaches to, evaluating the effectiveness and efficiency of service delivery in remote and discrete Aboriginal and Torres Strait Islander communities.*

*This chapter discusses these issues and makes recommendations to improve the culture and effectiveness of evaluation. It should be read in conjunction with the reform proposal outlined in Chapter 7.*

## Key points

- Evaluation is especially important in remote and discrete Aboriginal and Torres Strait Islander communities, since in many of these places, services are delivered or funded by government, giving residents little choice about the services they access, and reducing market incentives to improve effectiveness or efficiency.

- The reforms discussed in previous chapters should be evaluated, to keep them on track and to ensure they are leading to improved outcomes, including value for money.

- The existing Queensland Government Program Evaluation Guidelines provide sufficient guidance for conducting ex-ante evaluations. However, these ex-ante approaches, on their own, are unlikely to be sufficient to drive or support the level of innovation required to achieve better outcomes in remote and discrete communities and may even undermine progress.

- There are practical considerations that make evaluation difficult in remote communities, including difficulties establishing causality and the long timeframes required to achieve change. Even the most well-designed evaluations can leave themselves open to interpretation based on prior opinion, philosophy or politics.

- A more practical and informal evaluation approach with more relevant and timely information to communities will be most effective in developing solutions that work in a local context. Communities need to be involved in monitoring and evaluation, including the design of indicators of progress that are important to them.

- The structural reforms discussed in previous chapters will help to underpin and incentivise evaluation, since agreed expectations about outcomes, and how these are to be measured, should be set out in agreements between government and communities.

- The Commission also recommends the government adopt a framework for evaluation that:
  - assesses the progress of agreed reforms
  - reviews jurisdictional (regional or community) outcomes
  - assesses the extent to which localised decision-making and innovation is supported.

- Independent oversight of the framework will be required to ensure that stakeholders have confidence in the reform process and that evaluation is free from bias.

- Existing grant funding arrangements should be reviewed to ensure that compliance requirements are necessary and proportionate to the level of funding and risk.

# 10.1 Introduction

Evaluation is an essential component of service delivery, particularly where services are delivered or funded by government (Muir & Bennett 2014). It provides a systematic mechanism for collecting information and using this data to understand what is working and what is not. It can help stakeholders to track progress as programs are rolled out, and to determine whether a program is achieving its intended outcome, remains the best policy response and is delivering value for money:

> Evaluation matters. Too often it has been an afterthought … seen as an optional luxury for well-funded programmes, or done only if a donor requires it. This must now change, so that the role of evaluation is understood as an opportunity for organisational and individual learning, to improve performance and accountability for results, and to build our capacity for understanding why some programmes and initiatives work, and why others do not. (WHO 2013, p. v)

Evaluation is particularly important for service delivery in remote and discrete communities:

- In remote communities, mistakes may be hidden from general view—there may be significant, and costly lags before mistakes are identified and rectified, unless appropriate evaluation mechanisms are in place.

- Residents often have little choice when it comes to service provision as there is generally only a single government funded service provider—rigorous program evaluations help to make services accountable to users.

- Finally, evaluation, when done properly, can inform program design and help providers to adapt the way they deliver services to meet the needs of Aboriginal and Torres Strait Islander people in remote and discrete communities.

Despite widespread recognition its importance, the current approach to evaluation and monitoring of service delivery in remote and discrete communities is not working as well as it could. This is holding back the design and delivery of services, constraining innovation and failing to ensure that service delivery is accountable to the communities it should be serving.

To some extent, the deficiencies of evaluation can be applied to service delivery more broadly. Many of the same issues that apply to remote and discrete communities—such as a lack of funding for evaluation, a failure to build evaluation into program design and a lack of transparency—also apply to many government-provided or -funded service delivery across Australia. However, the lack of evaluation has disproportionately large impacts in remote and discrete Aboriginal and Torres Strait Islander communities, who are almost completely reliant on externally funded and delivered programs, and are far removed from decision-making processes that happen in agency head offices.

Changing this will be challenging. It requires cooperation across government, long-term commitment and need to provide room for communities and service providers to innovate and adapt to changing circumstance. There are also real technical challenges that make effective evaluation difficult in remote and discrete communities.

This chapter:

- provides background about evaluation—what it is, how it is done, what best practice looks like and the key challenges presented by remote and discrete communities

- makes the case for change—including stakeholder experiences

- presents a framework for evaluation of service delivery in remote and discrete communities.

# 10.2  Background to evaluation

## Typical approaches to evaluation

There are three commonly described approaches for program evaluation (Gertler et al. 2011; NSW Government 2016; Queensland Treasury 2014):

- **Process evaluation**—is the service implemented, is it being delivered and is it reaching the people as it was intended?

- **Effectiveness (or impact) evaluation**—is the service achieving the outcomes it intended to achieve?

- **Efficiency evaluation**—is the program efficient (are there better ways of providing the service) and does it provide value for money (do the benefits outweigh the costs)?

These methods typically rely on an understanding of the way in which inputs, activities and outputs are used to achieve outcomes—the program logic (Figure 53).

**Figure 53 Program logic—from needs to outcomes, evaluation can assess any or all of the components of a program as well as the outcomes the program is seeking**
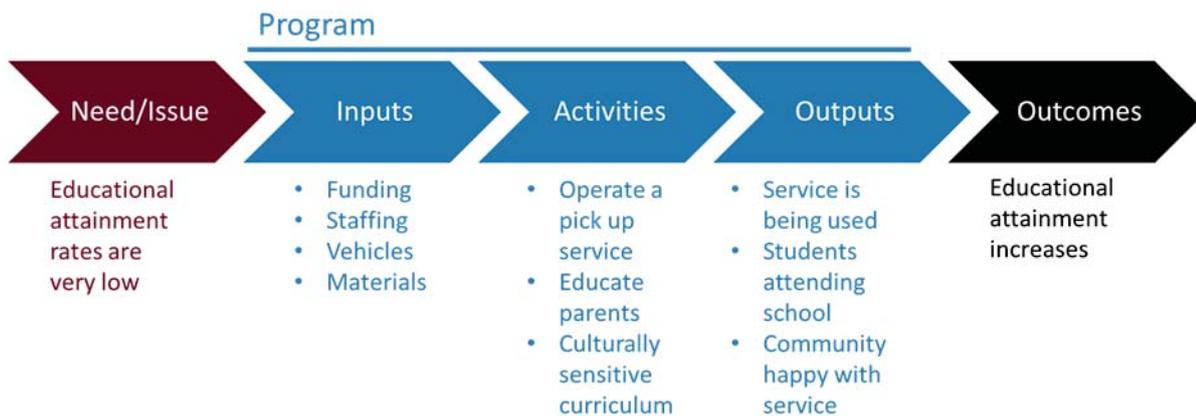


A **process evaluation** is the simplest of the three approaches and focuses on the inputs (staff, equipment, expenditures) and outputs (customers served, services provided). A process evaluation can assess the adequacy of service delivery mechanisms, management practices and the acceptability or suitability of services to stakeholders.

An **effectiveness (or impact) evaluation** focuses on outcomes. This type of evaluation is more complex, since it needs to establish a causal link between the program and any changes to outcomes—this requires consideration of what would have occurred as a result of the program, in comparison to what would have happened in the absence of the program (the counterfactual).

An **efficiency evaluation** provides information on the extent to which a program is efficient and provides value for money. Efficiency may relate to whether the program is being delivered at the lowest possible cost, the program is delivered to the area of greatest need, and whether the program has continued to improve over time (for example, by keeping up to date with technological advancements). An efficiency evaluation may also take a broader scope and attempt to determine if the program provides a positive social return (the benefits of the program outweigh the costs) and whether better returns could be generated by spending the money on other things.

The approach adopted will depend on the purpose of the evaluation, the level of complexity required, practical difficulties and the costs of conducting the evaluation. Queensland Government Program Evaluation Guidelines set out the government's preferred methods and approaches for evaluation (Box 10.1).

# Box 10.1 Queensland Government guidelines

The Queensland Government Program Evaluation Guidelines (Queensland Treasury 2014) outline the broad principles expected for evaluations of programs (including services) funded by the state, and set out standards for planning, implementing and managing these evaluations.

The guidelines take a practical approach to evaluation, and advise that the evaluation approach (process, effectiveness or efficiency) used should suit the circumstances, including the size of the program being evaluated, the risks involved and the stage of program delivery.

The guidelines are voluntary, and there is no formal mechanism for supporting organisations seeking to conduct evaluations of service delivery.

They do not provide any specific guidance for evaluations in an Indigenous or remote community setting.

**Figure 54 The evaluation process—Queensland Government program evaluation guidelines**



*Source: Queensland Treasury 2014.*

There has been a growing push for evaluations to focus on achieving improvements in outcomes for the people they service rather than simply focusing on outputs (Gertler et al. 2011). While there is a general consensus that measuring outcomes is desirable, there is some debate amongst practitioners as to whether this is practical, particularly in the context of Indigenous affairs (Altman et al. 2008; Cobb-Clark 2013; Guenther et al. 2009; Hudson 2017; James 2012, Moran 2016).

At a roundtable discussion on evaluation of Indigenous policies, hosted by the Productivity Commission in 2012, many participants felt that evaluation should identify what works and why, and that the continuation of program funding should be questioned if outcomes could not be conclusively demonstrated. Others, however, stated that this was unrealistic and that problems in the underlying system architecture that make it difficult to isolate a program's outcomes do not mean that a program is not worthwhile (PC 2012).

## Best practice

The World Health Organisation (WHO 2013) identifies five key principles for best practice evaluation:

- Impartiality—there should be an absence of bias in process which contributes to the credibility of the evaluation and allows findings to be accepted and trusted by stakeholders.

- Independence—evaluation should be free from the control or undue influence of others, including from policy makers and program participants.

- Utility—findings should be relevant and useful to stakeholders. Reports should be accessible to stakeholders, available for public access and there should be systematic follow-up of recommendations.

- Quality—evaluation criteria should be appropriate and accurate, evidence should be presented impartially and there should be coherence between findings, conclusions and recommendations.

- Transparency—stakeholders should be aware of the purpose and objectives of the evaluation, the methods of evaluation and the purpose to which the findings will be applied. This requires continuous consultation and involvement with stakeholders, and that reporting is complete and made public, unless publication is not in the interests of service recipients.

To be most effective, evaluation needs to be incorporated into service delivery, rather than occurring as a separate process, or even afterthought (James 2012). Best practice requires that planning for evaluation should start during program design, and where required, baseline data should be collected prior to implementation. After the program has commenced, there should be ongoing feedback between the evaluation and the program, to refine the collection of data, improve program design and to drive continual improvement (Chaney 2012).

Mayne (2010) and Stewart (2014) argue that for evaluation to be effective, a culture of evaluation needs to be established—this includes leadership by example, visible and consistent support for evaluation, decisions routinely informed by results, honest mistakes tolerated and learned from and evaluation being adequately resourced and funded. Mayne (2010) acknowledges the difficulties this typically presents for government:

> *Where this involves organizations led by ministers, a question not really addressed is how to get politicians to support an evaluative culture, especially given their frequent predilection for controls when things go wrong, and concerns about waste and mismanagement. Obviously, forms of education might help, as might the availability of empirical evidence when needing to defend programs. I suspect also that evidence that adequate controls are in place would be needed, as well as evidence on results … Building an evaluative culture at the political level remains uncharted territory! (Mayne 2010, p. 22)*

Cobb-Clark (2013) argues that increasing transparency and a much wider dissemination of results is central to raising the standards on program evaluation. She draws parallels with the successes achieved in the health sector from increased transparency and argues that a higher level of openness:

- increases pressure on evaluators to lift their game

- allows evaluations to be assessed against sound scientific principles so that judgements can be made about which to weight more heavily and which to ignore

- provides opportunities for informed debate on the issues facing Indigenous communities

- increases the chances for sound decision-making.

There is also a growing recognition that evaluation approaches are not one-size-fits-all and that several issues need to be considered when conducting evaluations in remote Indigenous settings. Best practice approaches for evaluation in an Indigenous perspective might include:

- involving local people in the design and implementation of policy—evaluation is just another step in policy design and implementation and should also involve the residents of Aboriginal and Torres Strait Islander communities (James 2012, see also Box 10.2)

- engaging local Indigenous researchers—the use of local people increases acceptance and trust, ensures surveys make sense in a local context, and gains access to the views of people who are not normally asked for their perspectives (Price et al. 2012)

- ensuring that evaluations are culturally competent—the culture-based assumptions of both those doing the evaluating and those being evaluated are fully understood and accounted for (Chouinard & Cousins 2007)

- ensuring that the outcomes of an evaluation are accessible and useful to communities and local organisations—effective evaluations should act as catalysts for instigating change by empowering communities (Price et al. 2012; Taylor 2006).

## Box 10.2 Participatory Action Research (PAR)

PAR approaches may offer opportunities for improving evaluation and including participants in designing solutions for complex problems. PAR is not a one-size-fits-all approach, but is built on the central tenet that all participants, including service users, should be engaged in every aspect of evaluation from defining the problem, gathering and analysing data, to preparing recommendations (McGarvey 2007).

Dudgeon et al. (2017) describe the use of PAR in 11 communities in remote communities in Western Australia as part of the National Empowerment Project (NEP). The NEP is an innovative Aboriginal-led project built around the use of PAR to give communities a voice, to design strategies for promoting well-being and building resilience, and for supporting change processes at both the individual and community level.

The authors suggest that the PAR approach played a key role in achieving real outcomes by building capacity and capability for individuals to take charge of their lives and support their communities.

> *This project gives a voice to communities in identifying the factors impacting on their social and emotional well-being and supporting them to see themselves as agents of social change. (Dudgeon et al. 2017, p. 10)*

## 10.3  Limitations and challenges for conducting evaluation in remote and discrete communities

The difficulties with measuring outcomes and/or attributing success (or failure) to a single intervention are accentuated in remote and discrete Aboriginal and Torres Strait Islander communities.

One reason for this is that many of the outcomes that program or services are trying to achieve may involve very long timeframes—for example, it has been estimated that it will take over 100 years to close many of the gaps in Indigenous disadvantage currently being targeted by federal and state initiatives (Altman et al. 2008). Although it may be possible to undertake a process evaluation (where the evaluation is concerned with measuring outputs, such as numbers of patients seen) early in the life of a program, understanding how a program affects outcomes may not be possible for many years.

A second reason is that it may be particularly difficult to establish causation in remote and discrete communities. Cobb-Clark (2013) identifies five factors that are specific to Indigenous communities that make evaluation of outcomes particularly problematic:

- For cultural, historic and political reasons Indigenous communities may be unique—this makes it challenging to define a meaningful control group against which to measure outcomes.

- It can be difficult to assess outcomes at the individual level because of extended, fluid family structures and cultural norms for resource-sharing.

- The highly political nature of Indigenous policy means it can be difficult to have a scientific process for random selection for treatment.

- The myriad of interventions likely to be occurring within communities means that it is difficult to single out a particular program or control for differences between communities.

- Many data sources have insufficient numbers of Indigenous respondents for analysis.

A further complication is that many organisations in remote and discrete communities have a broader mandate than is covered under their funding agreement. During consultation, we observed that almost every organisation we spoke with was undertaking a wide range of tasks, and in many cases, were spending most their time on non-core activities.  This generally was not by choice, but because many clients had immediate needs (such as a need to sort out issues with Centrelink) that were not being addressed elsewhere.

These issues suggest attempting to measure progress towards the achievement of outcomes can sometimes waste valuable resources and, at worst, may be used to justify funding being cut or programs modified because progress cannot be demonstrated (Chaney 2012). On the other hand, Guenther et al. (2009) disagree with an assessment that outcomes cannot be measured, arguing that complex programs in remote Indigenous settings can be assessed, but that the measures of success are not necessarily numerical or easily fit government reporting requirements.

Stakeholders expressed concern that the ambiguity inherent in even the most rigorously conducted evaluations carries the risk that they can be misused:

> [T]he answer to the problems that afflict our children, families and our places will not be found … even through the most conscientious application of the rhetoric about the need for more rigorous evaluation, application and adaption of 'what works' evidence or 'best practice' approaches. Indigenous communities are not laboratory environments, and the science of evaluation in such complex settings is not as precise as much of the rhetoric may suggest … in such cases, even impact evaluations using the best quasi experimental design, leave themselves open to still be used on the basis of opinion, philosophy or politics. (CYI sub. 26. p. 23)

A more pragmatic approach may well be required that acknowledges the process of by 'muddling along' and learning-by-doing, rather than following a prescriptive policy agenda (Chaney 2012; Moran 2016). Adopting this more pragmatic approach to evaluation may well mean that current evaluation approaches need to be adapted to allow greater flexibility.

# 10.4  The importance of data

The way that information and data is collected, used and presented can have real impacts on Aboriginal and Torres Strait Islander communities. Data are used to shape debate and influence policy decisions, identify areas of greatest need and direct resourcing, and allow monitoring of progress over time (Biddle 2014).

Data that are used well provide a robust evidence base that can help ensure that policies are well designed and that resourcing is going to the right things (Banks 2009). However, data collections that are poorly designed, poorly understood, incomplete or used incorrectly can distort decision-making and lead to inefficient or wrong outcomes.

The outcomes of interest to remote and discrete Aboriginal and Torres Strait Islander communities can be very different to the outcomes of interest to governments (Biddle 2014). This can lead to a disconnect between policy formation and the achievement of progress in improving community wellbeing on the ground. It is now widely accepted that effective policy and program design needs to be led by (or at least informed by) Indigenous people rather than government (PC 2012). Enabling this kind of community-led decision-making can only happen when communities have access to the kinds of information in a form that is useful to them (Biddle 2014).

It is widely recognised that Indigenous people's perceptions and understandings of wellbeing extend beyond, and often conflict with, many of the indicators currently adopted by reporting frameworks (ABS 2001; OECD 2011; Taylor 2006; Malezer 2012)). The United Nations' workshop on Indigenous Peoples and Indicators of Wellbeing concluded with a series of forceful statements on the need for more rights based indicators, including control over land and resources, equal participation in decision-making, preservation of culture and control over development processes (United Nations 2006).

Given the difficulties with attributing outcomes to any single service or program, and the need to consider Indigenous perspectives, it would be appropriate to measure progress against broader and more holistic measures of wellbeing than are being used in the current, limited reporting on outcomes in communities (Limerick 2009). Wellbeing measures have been widely used to track and report on progress in Indigenous communities around the world (Box 10.3).

## Box 10.3 Wellbeing Indicators

Wellbeing measures take a broader view of progress than typical economic indicators such as GDP or employment and usually combine a number of economic and social indicators into a single easily tracked measure (ABS 2001). They normally use a range of social indicators, such as health, social cohesion and economic status, that matter to the individuals or communities concerned.

A number of approaches have been adopted across different jurisdiction. However, most measures use relatively simple metrics that are collected regularly (such as through the Census or through regular social surveys). Some of these are considered below.

**OECD Wellbeing Indicators**—a compendium of indicators that the OECD considers as important measures of wellbeing (OECD 2011). Indicators include measures relating to material living conditions (income and wealth, jobs and earnings and housing) and quality of life (health status, work and life balance, education and skills, civic engagement and governance, social connections, environmental quality, personal security and subjective wellbeing).

**The Canadian Community Wellbeing Index (CWB)**—a measure of the standard of living and quality of life for all Canadian communities, including First Nations (Flanagan and Beauregard 2013). The index is produced by Indigenous and Northern Affairs Canada (INAC) and is composed of data on income, education, housing conditions and labour force activity.

**United Nation's Development Programme's Human Development Index (HDI)**—the HDI was created on the understanding that people and their capabilities should be the ultimate criteria for measuring the development of a country, not economic growth alone. It combines three key dimensions of human development: a long and healthy life, being knowledgeable and having a decent standard of living (Anand & Sen 1994).

**Social Progress Index (SPI)**—a measure of social progress developed by the non-profit sector as a response to the over-reliance on economic indicators. The index excludes any economic variables and is based on a range of social and environmental indicators based on three dimensions of social progress: basic human needs, foundations of wellbeing, and opportunity (Stern et al. 2016).

**Growth and Empowerment Measure (GEM)**—the GEM was developed as a tool to measure the progress that various interventions were having on empowerment. It seeks to measure people's perspectives of their psycho-social wellbeing and empowerment at the individual, family and organisational level (Haswell et al. 2010).

## 10.5  Current practice—assessment of effectiveness of evaluation in remote and discrete communities

The evidence available to this inquiry suggests that evaluation of service delivery in remote and discrete communities is not as effective as it should be. This was confirmed in our review of the literature (Box 10.4) and from conversations with stakeholders.

A common complaint is that there is little proper or effective evaluation of programs with (DOFD 2010; Hudson 2016b; SCRGSP 2016a). In Queensland, there are few publicly available, formal evaluations of service delivery to Indigenous communities. Those that are available tend to be part of larger Australian Government programs, such as the recent review of the National Partnership Agreement on Remote Indigenous Housing (DPMC 2017).

However, the problem is not simply one of 'not enough evaluation'. It appears that there is a significant amount of time, energy and money is spent on monitoring and evaluation in Indigenous affairs. For example, the Australian Institute of Health and Welfare's Closing the Gap Clearinghouse Research and Evaluation Register reveals 1,249 evaluation studies. Our tentative view is that the effort spent on evaluation in the Indigenous space is comparable to the effort spent evaluating social services, generally.

Stakeholders were clear that current evaluation efforts are not working.

> *The other thing that's really difficult in this is the evaluation stuff. And I don't have any easy answers for you here. I just know that the typical evaluations…conventional evaluation frameworks. They are of limited use…I can tell you that most of the evaluations that are done don't actually tell you what happened (Interview with Mark Moran)*

A big part of the problem is that current evaluation efforts focus on the wrong things. Most effort appears to be on either ensuring accountability for expenditures (compliance) or demonstrating value for money, rather than on finding ways to improve outcomes. This was reflected in stakeholder comments:

> *I would urge the Commission, however, not to advocate for a narrowly conceived approach to monitoring and evaluation that focuses only on service outputs and value for money measures. Rather, government's formulation of what are the "funded outputs" for services and programs in Indigenous communities needs to recognise that community capacity-building and community governance are significant outputs of government investment in themselves… Government funding models need to recognise and place a value on the capacity-building outputs, as it is this capacity that has the potential to create long-term positive change in Indigenous communities (Michael Limerick sub DR05, p. 5)*

> *Rather than ad-hoc program specific evaluations, regular (3-5 year) cycles of evaluation are needed that evaluate a package of programs directed to a specific problem or outcome. (PICC sub. 29, p. 10)*

> *Use of a reflective practice approach to evaluation relies on a two-way exchange, with the experiences of those delivering the program being used to inform its ongoing implementation. Although this approach might not meet the 'gold standard' in terms of research evidence, it would be more practical and achievable given limited resources. There is no point conducting 'rigorous' evaluations, if the evidence is not used. Instead of focusing on having the highest standard of evidence for assessing the impact of a program (such as in RCTs), it may be more practical to consider how to ensure evaluation learnings are used to inform program practice, similar to continuous quality improvement processes used in the health sector. (CIS sub. 21, p. 7)*

## Box 10.4 Evaluation in remote Indigenous settings—evidence from the literature

The literature suggests that evaluation efforts have not been as effective as they should be. This is not because evaluation cannot be effective, but rather, the system in which evaluation occurs does not allow for learnings to be taken on board and for service delivery to engage in adaptive practice.

> *The positive examples of evaluation which come to mind relate to corporations rather than government. I sit on the board of a provider which works in government schools, but with external funding coming from the private sector, mainly miners. We are regularly independently evaluated and have non-government funding relationships that extend over more than 15 years. Our funders are interested in learning as we go. They participate in the management of each individual project and, so far, have extended their support as we learn from what we are doing … This sort of evaluation leading to action makes sense.*

> *Miners, with their experience of project management, better understand taking responsibility for managing towards a long-term outcome, dealing with complexities and problems as they arise. You learn from mistakes, stop doing things that are unproductive and actively look for what will work. You do what needs to be done to achieve the objective. (Chaney 2012, pp. 60–61)*

This inability to encourage adaptive practice often appears to arise because of political demands:

> *Many – perhaps even all – public policy evaluations in Australia are conducted under exactly these [political] sort of constraints. However, while it may be possible to 'rescue' some semblance of an evaluation strategy with very clever lateral thinking, it is critical to realise that in the end we may not have actually learned very much. Often 'better than nothing' passes for 'good enough', leaving us as uninformed as ever, despite having spent millions (or even tens of millions) of dollars on the evaluation exercise. (Cobb-Clark 2013, p. 85)*

These political demands, in turn, encourage a cyclical pattern of change:

> *We have titled this report 'Don't let's lose another good idea' as a reflection of the concern that just as the evidence is emerging that something is working well, there will be a repeat of the old pattern of dispensing with a good initiative and trying something new. The review team urges that the lessons from this initiative are understood as being extremely relevant today and that the growing confidence of Indigenous communities in the SRA process is recognised and built upon. (Morgan Disney 2007, p. 7)*

A better approach may be for more informal approaches to evaluation:

> *The challenge then for Indigenous affairs policy is how to take the current ad-hoc standards of 'muddling' or 'gaming' in practice, to a more proactive and transparent alignment with policy. This would require administrative mechanisms to be built around local organisations rather than external grants.*

> *The nature of Indigenous development is flawed and fluid. Mistakes will be made and resources will be captured by elites, much as it occurs in mainstream local government. There will be capability gaps and, at times, corruption. But if the systems are designed around these weaknesses alone, then the result will be less space for capacity and innovation to grow. (Moran 2016, p. 195)*

Perhaps most worryingly, we found little evidence of evaluation being used to support continuous improvement or innovation in service delivery. Rather, most monitoring and reporting appeared to be compliance related, generating little useful information for communities or service providers:

*There is little evidence of either measurement or evaluation of activities being undertaken by stakeholders in the region to determine whether these are having a real impact on the level of disadvantage being experienced or demonstrating a return on investment for funding bodies. (TRSA sub. 22, p. 3)*

*The reporting requirements are particularly unhelpful for organisations such as PICC as they do not capture the broad nature of the work and outputs/outcomes achieved as a consequence of the organisation working holistically within the community. (PICC sub. 29, p. 9)*

*Reporting requirements which focus on output reporting, do not provide the information required to report on the outcomes or longer-term impact achieved by an organisation. Additionally, reporting requirements which focus on capturing the number of minutes spent with a client may account for 'Service Hour' reporting, but does not account for the quality and level of a service provided. (ICAN sub. 17, p. 12)*

*At times it is clear that funding bodies place more emphasis on funding compliance than actual service delivery outcomes. Often compliance issues raised with the organisation are of a minor nature such as a query over the percentage allocation of funding to individual cost codes within a service budget. Responding to such queries is time consuming and repetitive, and the requests are often generated by Departmental staff with limited understanding of the operations of an NGO (PICC sub. 29, p. 8)*

*What is required is 360 degree evaluation. However, the reality is, government doesn't want to be evaluated. (Stakeholder meeting, Cherbourg)*

*Torres Strait Island Regional Council (TSIRC) has 15 sets of infrastructure to supply drinking water to its 15 communities. Drinking Water Quality Management Plans are required for each individual system. The result is that, while a small remote council of 5,000 residents, the TSIRC is completing almost eight times the reporting of large well-resourced councils such as Cairns Regional Council.*

*TSIRC find that this reporting is not of much benefit, as it is too cumbersome to keep up to date and does not add value to service provision or asset management. The limited staff and engineering consultants are spending valuable time reviewing, auditing and updating drinking water quality management plans which could be better spent solving problems with operations and writing grant applications for much-needed funds.*

*Because TSIRC faces fines if these reports are not submitted in a timely manner, it is forced to allocate resources towards completing them. This in turn means that there is not human resources focused on resolving the issues on the ground. We then need to hire engineering consultants to assist in completing compliance reports, at considerable cost.*

*A shift in the approach to compliance in Departments such as DEWS and DEHP is essential. The one-size-fits-all, inflexible and punitive response does not work to achieve outcomes for TSIRC's communities and it is not solution-oriented. (TSIRC sub. 12, p. 39-40)*

The amount of effort spent on monitoring and evaluation appears to be disproportionate to the level of risk. Stakeholders told us that there is little relationship between the level of risk or funding and reporting requirements. Many stakeholders told us that they spend enormous amounts of time and effort on reporting for little obvious reason.

> *Within departments there can be wide variation in the onerousness of grant conditions; with one being a low burden on Council and another high. … current flexibility within Departments to set grant conditions is resulting in inconsistent processes at the delivery level for Council. We are in a position with some grants, such as the Department of Communities, $18,818 HACC grant where the cost and resources required to administer the grant makes the service delivery on the ground almost negligible … Consideration should be given to scaling levels of reporting dependent on recipient's financial standing. (TSIRC sub. 12, pp. 17-18)*

The requirement to source grant funding from multiple sources adds large compliance burdens particularly when funding must be sourced from two levels of government, each with different reporting requirements.

> *Where dual funding occurs across State and Federal programs, it would produce far better outcomes on the ground if the State and Federal governments invested up-front in working together to develop and streamline reporting requirements. The failure to do so has significant impacts on productivity and service delivery on the ground. (TRIRC sub. 12, p. 39)*

The current system also lacks transparency. Stakeholders raised concerns about the lack of information relating to large programs, including alcohol management programs and Welfare Reforms, however, they also cited concerns about the general lack of information and data available to communities, service providers and local decision-makers.

> *There is very little evidence available around program evaluation and reporting. When asked, service providers say that "the council or community are not entitled to access this information – that it is confidential" and that they are only required to provide this to their funding agency. Government agencies have also been reluctant in the past to provide information about funding levels or real time data on program progress. (LGAQ sub. 14, p. 40)*

The lack of transparency (whether this is perceived or real) has generated a general mistrust of government and the way that services and programs are evaluated. As a result, many stakeholders wanted more independent oversight of evaluation and data. For example:

> *A central independent body to coordinate and oversee evaluation in Aboriginal and Torres Strait Islander communities is an interesting suggestion. If done well, this may reduce the evaluation burden on communities (by identifying and avoiding duplication), build Aboriginal and Torres Strait Islander Evaluation capacity, and deliver on promises to co-design evaluations with community and provide feedback about findings. It could also be a mechanism to continuously collect outcome data common to many services and programs, and minimise the data required for individual evaluations. (Anon.)*

> *When external evaluators are used, it is common for the government to insist that the results not be published. In short, the results of these evaluations are typically not independent, transparent or widely distributed. … All of this is inconsistent with the move to evidence-based policy and undermines our ability to deliver on closing the Indigenous gap, raising educational achievement, and reducing social exclusion. (Cobb-Clark 2013, p. 90)*

The TRSA went one step further suggesting there needs to be independent assessment of reporting and outcomes:

> *There is also a need to establish an Indigenous Ombudsman Office in Queensland to ensure there is an independent umpire to make sure targets and outcomes are met by all stakeholders delivering services in the region. (TRSA sub. 22, p.3)*

# 10.6 The scale of evaluation should suit the scale of reforms

Government service delivery has a large impact on remote and discrete communities. It cuts across multiple agencies, all levels of government, involves large sums of money and directly impacts on the service recipient behaviours (such as through welfare dependency, mobility and self-responsibility).

While there are numerous individual programs and service providers, their impacts of each are co-dependent. This means that it is virtually impossible to establish causality between any single program and outcomes observed in communities.

The reforms proposed in this report are also large and involve long time frames. A significant evaluation effort is needed to keep them on track.

The Queensland Government Program Evaluation Guidelines suggest that the scale of evaluation needs to suit the scale of the program (Table 27). The guidelines imply that any evaluation plan will need to allow for:

- an assessment of implementation of reforms

- an evaluation of impacts on outcomes

- independent oversight

- extensive consultation.

**Table 27 Scaling evaluation - Queensland Government Program Evaluation Guidelines**

| Program characteristics | Evaluation design | Evaluator | Stakeholder consultation |
|---|---|---|---|
| Low risk<br>Simple program design<br>Low resource requirements<br>Ongoing program<br>Single delivery body<br>Low potential for behavioural impacts | Qualitative assessment of implementation success and program efficiency built into program reporting could be sufficient.<br>Additional assessment should be undertaken where it is cost effective to do so. | The evaluation can be conducted by an internal evaluator, seeking advice and assistance from experts if required. | Consultation may simply involve program managers, but could also include individuals or institutions directly impacted by the program. |
| High risk<br>Complex program design<br>Pilot or trial program<br>Multiple delivery bodies<br>High potential for behavioural impacts | Comprehensive evaluation design that assessed the programs' implementation, efficiency and effectiveness from a whole of society perspective. | The evaluation should be managed by an external party, independent of the agencies involved in program delivery. | Extensive consultation with both the stakeholders listed above as well as:<br>- Evaluation experts on the appropriateness of evaluation design, methods and assumptions<br>- Central agencies on the fit-for-purpose nature of the evaluation approach and potential findings. |

*Source: Queensland Treasury 2014.*

## 10.7  A framework for evaluation in remote and discrete communities

This section lays out an approach for evaluation in remote and discrete communities. It provides a framework that provides a balance between practical considerations, required accountability measures and the need to arm decision makers, service providers and program designers with the tools they need to improve on-the-ground service delivery.

There are three elements to the proposed framework:

- to keep reforms on track—an evaluation of the implementation of structural reforms based using a process evaluation approach

- to assess whether outcomes are getting better—a jurisdictional-level assessment of outcomes relative to a counter-factual (business-as-usual outcomes)

- to enable effective service delivery—support for individual program/service evaluation with a focus on enabling adaptive learning.

To provide confidence to all parties, the evaluation framework will need to be underpinned by independent oversight.

**Figure 55 A proposed evaluation framework**
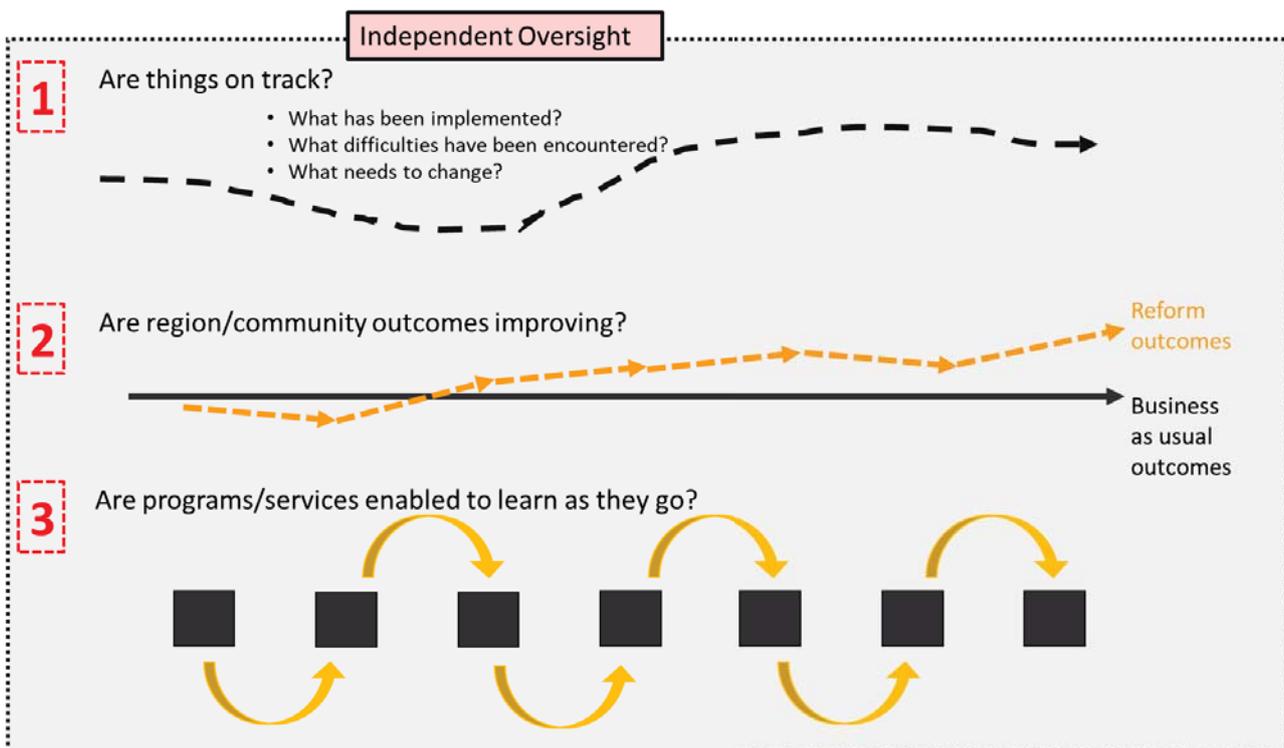


The framework provides practical evaluation strategies that help to overcome many of the challenges identified earlier in this chapter. It will help ensure reforms are being implemented in the best way and remain on track, have the best chance of delivering real improvements for remote and discrete communities and are delivering value for money on public investments. Done well, the evaluation strategy should support improvements in service delivery and engender broader support for long terms reforms to the service delivery environment.

# Element 1: Evaluation of reform progress

The first element the recommended evaluation framework is to assess whether reforms have been implemented as agreed and whether outputs are occurring as intended.

The purpose is to understand what has occurred to date and to learn lessons that will improve implementation moving forward. The focus should be on inputs, outputs and processes, *not* outcomes.

For this stage of the evaluation it is important to understand the logic behind the reforms, and to ensure that the evaluation focusses on the key outputs that were required to achieve change (Figure 56). For example, the logic behind the structural reforms in Chapter 7 is that agreement making to devolve decision making and authority to communities will improve service delivery and outcomes. This means that a key focus of the evaluation should be to assess the agreement making process—this requires an understanding of the inputs used to support the process (and whether they were sufficient), what activities took place and whether outputs were produced as intended.

**Figure 56 Process evaluation of reforms requires an understanding of the program logic**
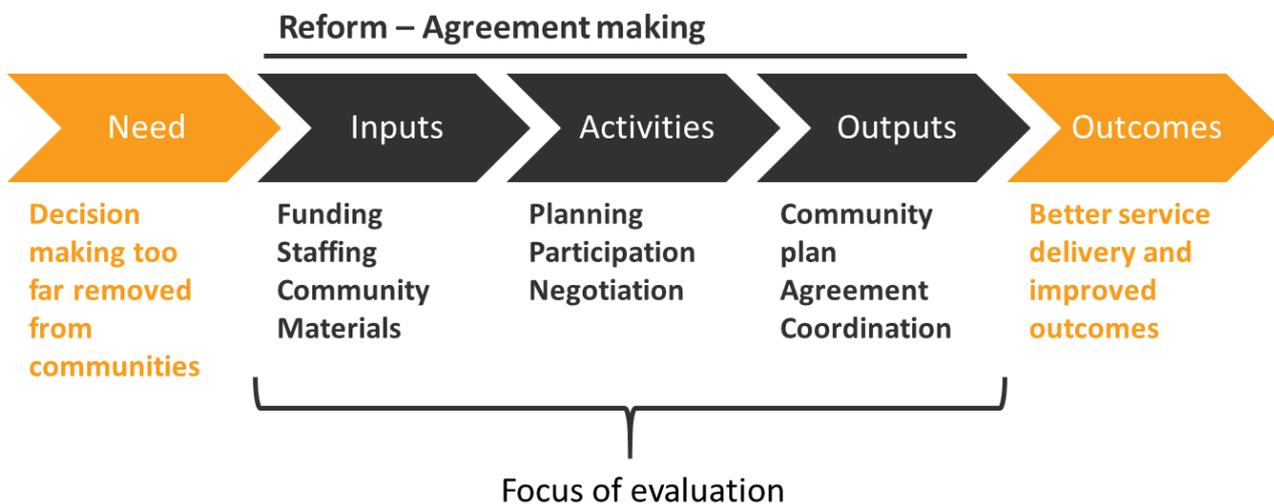


While many factors will determine whether the reforms have been implemented successfully, key components will include that the reforms have been implemented on schedule and to budget, have delivered the agreed outputs to a sufficient standard and have met the expectations of stakeholders.

The key questions the evaluation should ask are:

- What has been implemented and is this consistent with the intent of the reforms?

- Have agreements been made and adhered to?

- What is the quality of governance arrangements?

- Have agreed milestones been met and targets for outputs been achieved?

- Are outputs of a sufficient quality and are they likely to influence outcomes as intended?

- Are stakeholders engaged with the reform process and what can be improved? Are stakeholders satisfied and do they have legitimate concerns?

- What lessons can be learned about implementation so far, and what needs to change?

The intent of the evaluation should be to understand what has happened during the reform process and why. It should identify any unintended consequences that have occurred during implementation and seek to understand how the implementation of reforms can be improved and built on.

Qualitative and quantitative evidence will be required to support this element of the evaluation framework.

It will be particularly important to gather stakeholder views on the integrity of the process and how things can be improved moving forward. The literature suggests that it will be important to enlist Indigenous evaluators in gathering the views of Aboriginal and Torres Strait Islander people living in communities can be heard (Price et al. 2012).

Evaluators will also need access to quantitative evidence, including expenditure and staffing levels, any monitoring data (for example, participation at community meetings) and program reports.

Consideration for how this evidence will be collected and collated should be made during the implementation stage of any reforms adopted by government.

## Element 2: Regional and community outcomes

To ensure accountability, and to provide confidence that large expenditures are achieving meaningful change, there needs to be monitoring of the outcomes occurring in communities and/or regions.

Given the enormous challenges in attempting to establish causal links between individual services or programs and community-level outcomes should often be considered the result of the service delivery system as a whole, rather than linked to any single program or service. Recognising that inter-relationships between the various programs and service providers is crucial to achieving outcomes means that meaningful results will be achieved by looking at outcomes at a region or community level rather than through service silos.

When looking at outcomes, it will be important to understand the jurisdiction that authorises service delivery. Under the proposed structural reforms, decisions about service delivery are made at either community or region level jurisdictions. This means that the appropriate regional level for examining outcomes may vary depending on the outcome being examined. For example, it would make little sense to examine community employment rates where regional approaches are being used to create employment by encouraging mobility. Similarly, it would not be helpful to examine outcomes at a regional level where different service delivery decisions are being made by each community in the region.

The outcomes to be included in any evaluation framework also need to consider the perspectives of Aboriginal and Torres Strait Islander people living in remote and discrete communities, as well as government and the broader community. These should be negotiated as part of the agreements between communities and government, and should consider holistic measures of well-being.

The Australian Bureau of Statistics (ABS) National Aboriginal and Torres Strait Islander Social Survey covers a range of social indicators that could form a more comprehensive picture of community wellbeing. As discussed in Chapter 7, consideration should be given to working with the ABS to extend the sample size of this survey, to provide a more complete picture of progress in remote and discrete communities.

When monitoring outcomes over time, it is important to consider how these change relative to a baseline (or business as usual expectations). Establishing these baseline expectations is necessary if decision makers are to make realistic assessments of the performance of the service delivery system (Box 10.5).

## Box 10.5 The importance of establishing baseline expectations

As shown in Figure 57 the interpretation of success is heavily influenced by performance relative to the baseline. For example, while outcomes may improve in absolute terms, reforms should only be considered a success if they improve things relative to the usual way of doing things. If the pace of improvement slows, any objective observer should conclude that reforms have not been successful. Similarly, if reforms do not lead to improvements in absolute terms, but stop a previous decline in outcomes, they should be considered successful.

**Figure 57 Outcomes – three possible scenarios**
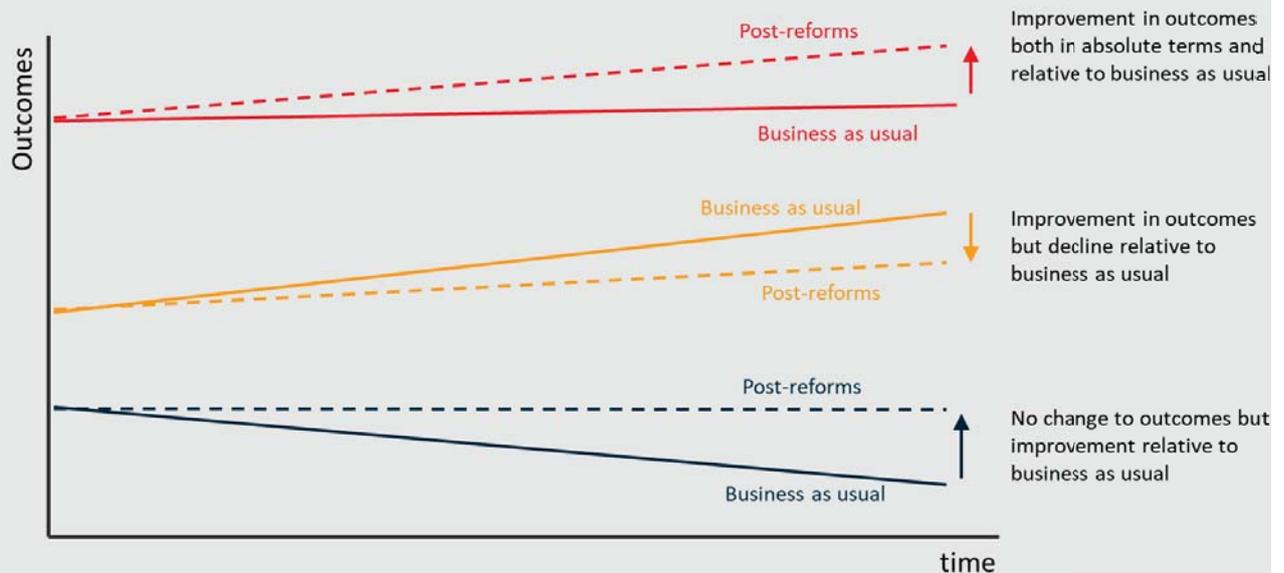


Establishing baseline expectations about outcomes will not be easy. There are a wide range of factors to consider, including historical trends, but also endogenous factors that may influence outcomes. For example, when establishing a baseline for employment outcomes, economic influences such as mining investment activity would need to be considered as they have a large influence on the ability of service providers to influence employment outcomes in remote areas.

In the absence of any other information, the Commission recommends that business as usual outcomes be assumed to show no change. That is, the default assumption should be that the current service delivery system is having little effect on outcomes.

The other important consideration to make when assessing performance is the level of resourcing that accompanies any reforms. If there is an increase (or decrease) in resourcing to communities under the reform process, then expectations regarding performance should also increase (or decrease). Again, determining whether resources to communities have increased or decreased is not straight forward and may need to consider whether reforms have changed the level of resourcing that hits the ground in communities, rather than just the absolute levels of expenditures made across the service delivery system.

# Element 3: Evaluation to support program/service delivery

Perhaps the most important consideration of any evaluation framework is how well it supports improvements to service delivery.

A range of practical difficulties make evaluation challenging, not the least of which is that evaluations are expensive. Hudson (2017) estimates that the average cost of an evaluation is $382,000. Clearly it will never be practical to conduct evaluations of every program implemented in Aboriginal and Torres Strait Islander communities.

Another factor that makes conventional evaluation approaches difficult to implement is that the potential solutions to the complex problems facing remote and discrete communities are not always evident. Typical evaluation requires a program logic (a map of how program inputs and outputs will lead to change). However, this can be impossible to establish in absence of known potential solutions.

Service or program evaluations are also made more difficult by the fact that the activities of funded organisations are rarely constrained to their funded activities.

These factors mean that a standard approach to program evaluation is often unlikely to provide useful information to practitioners dealing with complex problems in communities.

If evaluation is to better support service delivery improvements it must move away from the traditional approach, typically imposed from the top down, which requires strict accountability requirements, but fails to recognise there may be better ways of doing things.

The evaluation framework must enable and foster local decision-making, and encourage adaptive practice (including learning from failure). For this to occur, evaluation, including reporting and compliance activities, must support the needs of communities rather than just government. This is not to say that an evaluation framework does not need to provide accountability for taxpayer's money—it does—but there should be greater consideration of how evaluation can improve information provision to the communities, individuals and stakeholders that are best placed to make day-to-day decisions about how services are delivered.

More recent literature suggests that a development approach to evaluation may enable an adaptive learning approach to service delivery in remote and discrete communities (Hudson 2017, Gamble 2009, Chaney 2013). Under this approach, it is recognised that developing solutions to complex problems is not a linear process. Rather, solutions are developed through incremental steps, each of which is an experiment—the evaluation approach needed to support this process is one that provides fast feedback and informs learning (Box 10.6).

> *There needs to be a recognition that conventional evaluation frameworks are of little use. What is required is an approach that allows true adaptive practice - this means real-time information with a focus on incremental change. This may mean the reality is that program logic and evaluation is done "on the fly". Evaluators need to get out of an "ex ante" focus - it doesn't work (interview with Mark Moran).*

## Box 10.6 Developmental approaches to evaluation

Development approaches to evaluation are relatively new and have grown out of observations on how innovation occurs in the corporate sector. It recognises that entrepreneurial logic is based around the idea that answers are not known and will only be discovered through a process of 'trying things' (Gamble 2009).

The approach is suitable where there are complex or 'wicked' problems that do not have obvious optimal solutions or occur within stable constraints. In these situations, it is not possible to take a linear approach to solution making where logical steps can be mapped out in advance. Rather, solutions must be developed with little prior knowledge about which approaches are most likely to work and are formed through experimentation, exploration and innovation.

A development evaluation approach is used to support adaptive practice and innovation (Hudson 2017). It is not tied to any particular stage in the roll out or delivery of a program. Rather evaluation is an ongoing effort which informs planning and action in incremental stages (Figure 58).

**Figure 58 Approaches to evaluation**



the traditional approach:
Plan → Act → Evaluate

the developmental approach:
Plan
Act
Evaluate
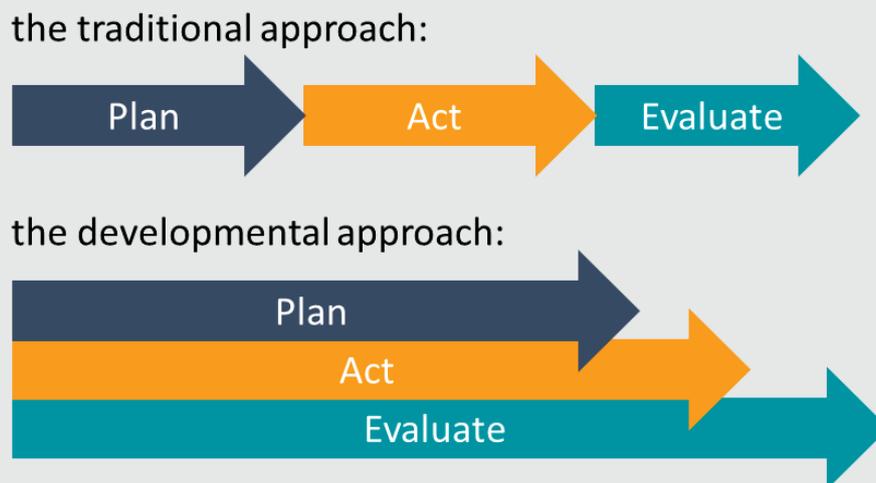
*Source: Adapted from Gamble 2009.*

Development evaluation relies the experiences of those on the ground delivering the program to inform the ongoing implementation of the program. It relies on real-time data—implementers need to be able to quickly see if what they are doing is having an effect and adapt their approach accordingly.

Development evaluations are not incompatible with other, more formal approaches, but recognise that these formal approaches can stifle the innovation required to improve outcomes.

Enabling a developmental approach requires that local decision makers have access to information. Currently, communities have access to very little information, and reporting and compliance efforts are almost solely directed towards government. Community stakeholders have told us they find it hard to find even the most basic of information—such as which services have been funded in their communities and what they are supposed to be delivering—let alone information on the performance of these services.

Any new model for compliance and monitoring as envisaged in the Commission's proposed evaluation framework must include more provision of information to communities than exists today.

Currently, service providers report to and are monitored by government. The focus needs to shift to the individuals and communities using services. The reforms proposed in Chapter 7 will achieve this outcome for the participating communities. For other communities, there must be a much greater commitment from government to provide information to communities. This should occur wherever possible using commissioning models that put communities at the centre of service delivery (as outlined in Chapter 8). Reporting requirements should be negotiated with communities and be included into service contracts with service providers.

Existing data, including non-confidential agency information, need to be readily available to communities and other stakeholders in a form that is timely, appropriate and useful. To support this, a commitment should be made to:

- regular reporting of outcome and expenditure information to communities—where the information needs of communities are negotiated with government and service providers.

- periodic, public release of information on progress and expenditures in remote and discrete Aboriginal and Torres Strait Islander communities.

Less formal evaluation approaches, such as development evaluation methods, should be used in combination with more formal or traditional evaluation methods.

Formal evaluation methods are still important and can provide useful insights and lessons about why programs do and do not work, and will continue to provide an important accountability measure for larger programs.  The crucial point, however, is that traditional evaluation approaches should be used only when they are likely to provide useful and meaningful information that can be used to improve service delivery.  Their findings need to be shared. Failure to do so feeds stakeholder perceptions that evaluation is little more than a compliance tool.

To this end, government should commit to:

- working with all stakeholders to develop evaluation strategies for any new major reform or program, including how the results of evaluation will be funded and shared with stakeholders

- the timely and public release of all evaluation reports for services it funds, either directly or indirectly.

# 10.8 Independent oversight

A common complaint is that the learnings from evaluations are ignored and that decisions to shift or change funding are made before evaluation have been completed (Morgan Disney 2007). If these complaints are to be overcome, evaluations must be de-politicised and used to support and identify good service delivery practices. This will require an approach that is at arms-length from government.

The Commission recommends that an independent body be tasked with overseeing the evaluation framework. The independent body should be responsible for:

- reporting the progress of reforms by conducting or commissioning a full review of reform implementation

- establishing baseline estimates for key indicators of interest to communities and government

- consolidating outcome data and reporting on these relative to established baseline estimates

- reporting on progress on outcomes, including an assessment of the performance of the service delivery system

- establishing the extent to which the service delivery system is supporting providers to develop innovative, locally driven solutions, and how information provision and evaluation effort can be improved to better encourage effective service provision

- acting as a clearinghouse for evaluations of individual services or programs.

The independent body should publish a report every two years outlining its findings.

The roles of the independent body could be performed by an existing body, or by establishing a new statutory authority. Regardless of who performs these roles, the body must have:

- a clear mandate, established in legislation

- indigenous representation, such as an Indigenous Commissioner

- the ability to operate at arms-length from the government

- sufficient resourcing and expertise to carry out its functions

- enough longevity to see reforms through to establishment.

# Recommendation 10

The Queensland Government should establish baseline estimates of expenditures made in each of the discrete communities, and for remote regions. This should include the proportion of expenditure spent on indirect or ancillary functions.

Actual expenditures in each of the discrete communities and remote regions, including indirect or ancillary expenditures, should be estimated every two years and made publicly available to support transparency and decision-making.

# Recommendation 11

The Queensland Government should commit to an evaluation and reporting framework that supports adaptive practice, facilitates accountability and empowers communities by providing them with timely, useful and relevant information. To support this framework, the government should:

- identify the outcomes communities are interested in tracking

- improve the availability of agency and other data to support local decision-making

- work with the Australian Bureau of Statistics to extend the coverage of existing surveys such as the National Aboriginal and Torres Strait Islander Social Survey

- develop evaluation strategies with communities for any new large reforms or programs, including how the findings will be shared with stakeholders

- make evaluation reports for services it funds, either directly or indirectly, publicly available in a timely way

- ensure that existing compliance requirements are necessary and proportionate to the level of funding and risk

- streamline reporting and compliance requirements for areas of shared responsibility with the Australian Government.

The Queensland Government Statistician's Office should regularly collate and provide data to communities to support the evaluation framework.

# Recommendation 12

The Queensland Government should assign an independent body to oversee and report on the operation of the Government's evaluation framework for remote and discrete communities. The functions of the independent body should include regular, public monitoring and reporting on:

- progress of reforms to the service delivery system

- performance against the agreements between communities and government

- outcomes being achieved in communities, relative to established baseline estimates

- the extent to which compliance, monitoring and evaluation efforts are supporting innovation and improvements in service delivery.

These functions may be allocated to an existing organisation but should be established in legislation, and include appropriate expertise and Indigenous representation.