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. After assessing the evidence, the Commission has concluded that existing frameworks, such as the Queensland Government’s program evaluation guidelines, provide sufficient guidance for evaluation. Rather than a new evaluation framework, what is required is a change in the service delivery framework that will better support effective evaluation.

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 many of these communities are dominated by government-funded service delivery, giving residents little choice about the services they access, and reducing market incentives to improve effectiveness or efficiency.

- Existing Queensland Government frameworks provide sufficient guidance for conducting evaluation.

- Compliance and reporting arrangements could be more open and effective. Many grants have high compliance costs that are unrelated to risk or the quantum of grant monies provided.

- Evaluation is rarely a part of program commissioning, tends to focus on outputs rather than outcomes, and does not appear to support adaptive practice—in some cases, evaluation has lacked sufficient transparency to enable communities and service providers to engage in adaptive practice.

- Many evaluations focus on a single aspect of service delivery and could add more value—holistic interventions are required to solve the complex issues affecting remote and discrete Aboriginal and Torres Strait Islander communities and this should be reflected in the evaluation design.

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

- 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.

- Communities need to be involved in monitoring and evaluation, including the design of indicators of progress that are important to them.

- Regular, transparent information on progress, expenditures and evaluation of service delivery is required to encourage adaptive practice and a service delivery model that provides value for money.

- Monitoring and evaluation need to be underpinned by independent oversight.
10.1 Why evaluate?

Evaluation is an essential component of service delivery, particularly where services are delivered 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 to assist stakeholders to track progress as programs are rolled out, to determine if a program is achieving the outcomes it was designed to accomplish, if it remains the best policy response and whether the program 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)

There are a number of reasons why evaluation may be especially important in the context of service delivery in remote and discrete Aboriginal and Torres Strait Islander communities.

The high cost of service delivery, combined with relatively higher needs of Aboriginal and Torres Strait Islanders living in remote regions, mean that mistakes in service delivery can be costly. This is likely to be exacerbated by the fact that service delivery occurs in remote areas where mistakes may be hidden from general view—as a result, there may be significant, and costly lags before mistakes are identified and rectified, unless appropriate evaluation mechanisms are in place.

Evaluation can help to ensure accountability to both the users and funders (the taxpayer) of services delivered in communities.

Importantly, in many remote Aboriginal and Torres Strait Islander communities, residents have little choice when it comes to service provision as there is generally only a single government funded service provider. In the absence of user choice, there is little incentive for service providers to achieve high levels of performance or keep up with the latest developments. In this case, rigorous program evaluations may provide an important mechanism for ensuring services are accountable to users.

Governments spend large sums of money addressing the high levels of disadvantage in Queensland’s remote and discrete communities. To ensure accountability to taxpayers, independent evaluations can provide an important mechanism for ensuring that this money is being used well and for the purposes it was intended.

10.2 Methods for evaluating

Evaluation can be undertaken in a number of forms, with varying levels of complexity and expense.

Three commonly described approaches for program evaluation are discussed in government guidelines (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)?
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.

In many cases a process evaluation may provide sufficient evidence for success, particularly where a course of action has been shown to lead to successful outcomes elsewhere. For example, a process evaluation may enable a program manager to adapt a successful program to local circumstances.

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.

There has been a growing push for evaluations to focus on the achievement of 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).

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 it not worthwhile (PC 2012).
10.3 Best practice in program evaluation

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

- Impartiality—the absence of bias in process. Impartiality contributes to the credibility of the evaluation and allows the formation of credible findings.

- Independence—freedom from the control or undue influence of others, including from policy makers and program participants. This should not exclude program participants, as they have an important role to play in any evaluation. Rather, evaluators must be free to draw their own conclusions from the evidence available to them.

- Utility—the requirement that findings are relevant and useful to stakeholders. This includes that reports should be accessible to stakeholders, available for public access and that there is systematic follow-up of recommendations.

- Quality—the appropriate and accurate use of evaluation criteria, the impartial presentation and use of evidence and coherence between findings, conclusions and recommendations.

- Transparency—the requirement that stakeholders are aware of the purpose and objectives of the evaluation, the methods of evaluation and the purpose to which the findings will be applied. In practical terms, this means that the evaluation should ensure continuous consultation and involvement with stakeholders, that reporting is complete and made public, unless it is not in the interests of service recipients.

Deborah Cobb-Clark from the Melbourne Institute of Applied Economic and Social Research argues that increasing transparency and a much wider dissemination of results is central to raising the standards on program evaluation (Cobb-Clark 2013). 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.

The literature suggests evaluation needs to be incorporated into programs, rather than occurring as a separate process, or even afterthought (James 2012). Best practice requires that planning 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) (see Box 10.1).

Mayne (2010) and Stewart (2014) argue further 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)

Box 10.1 Queensland Government guidelines

The Queensland Government maintains guidelines for program evaluation (Queensland Treasury 2014). These guidelines outline the broad principles expected for evaluations for programs (including services) funded by the state, and set out standards that should be used 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 53 The evaluation process—Queensland Government program evaluation guidelines

There is a growing recognition that evaluation approaches are not one-size-fits-all and that a number of issues need to be considered when conducting evaluations in remote Indigenous settings (Figure 54). 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)

- 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).

**Figure 54 The four R’s framework for research with Indigenous communities**

**Respect**
- Respect for the richness and integrity of Indigenous communities
- Respect for the cultural rights of Indigenous peoples
- Respect for cultural diversity

**Relevance**
- Ensuring evaluations inform social policy and guide good practice
- Consultation, negotiation and mutual understanding developed with those affected by the research

**Reciprocity**
- Equitable benefits for participating communities
- Use of, and access to research results and outcomes that are of benefit to communities

**Responsibility**
- Negotiation around the research process and informed consent
- Acknowledgment of Indigenous experiences
- Commitment to accountability and transparency

*Source: Adapted from Markiewicz 2012.*
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 an important role in achieving real outcomes through 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)

Hudson (2017, p. 19) provides some cautionary advice on PAR, citing research showing that, although involving local people can have positive impacts, it can also result in programs being controlled by local elites, which means more disadvantaged members of the community miss out.
10.4 Limitations and challenges

Data collections are never perfect

The way that information and data are 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.

As discussed in Chapter 4, publicly available data on Queensland’s remote and discrete Aboriginal and Torres Strait Islander communities provide only a partial picture of the outcomes occurring in these communities, and how these have changed over time. This means that those influencing decision making (including the media and general public who may influence the political process) do not have access to the complete story about the progress being made in discrete communities.

This may be exacerbated if the reported outcome measures do not represent the needs or wishes of Aboriginal and Torres Strait Islander people, increasing the chances that unwanted interventions may be imposed onto communities.

Data or information is often unavailable to stakeholders

Existing data are often difficult to access. This may occur because evaluation can threaten programs and personnel—as a result, affected people may seek to protect themselves by limiting or denying access to data (James 2012). Similarly, agencies may be loath to release data for fear of political ramifications.

Even where there are no deliberate attempts to obstruct access to data, there can be difficulties for third parties to gain access to administrative data, either because the data are not collated centrally (for example, where expenditures are spread across multiple agencies) or approval processes cause excessive delays.

Where data are held centrally, often custodians are not free to release this data to others (including to other departments) without the agreement of the agency custodians. This can create lengthy delays for evaluators and may have an impact on the quality of the evaluation.

Little data or information is in the hands of Aboriginal and Torres Strait Islander people

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). As discussed in previous chapters, communities have access to limited data, and have little control over the collection, use and presentation of data.
Evaluations may not provide clear answers

It should be acknowledged that there are real difficulties with measuring outcomes and/or attributing success (or failure) to a single intervention. While the same issues affect all evaluation, they 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 some form of process evaluation (where the evaluation is concerned with measuring outputs, such as numbers of patients seen) early on in the life of a program, understanding how a program effects outcomes may not be possible within the timeframes demanded by the political cycle.

A second reason is that it may be 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 at a given point in time means that it is difficult to single out a particular program or control for differences between communities.
- Many data sources are unsuitable as they do not have sufficient numbers of Indigenous respondents for analysis.

These issues may mean that in many cases attempting to measure progress towards the achievement of outcomes would be counterproductive. At best it could waste valuable resources and, at worst, may lead to funding being cut or programs modified because of a perceived lack of progress (Chaney 2012). 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 approaches need to be adapted to allow greater flexibility of service delivery approaches.
Box 10.3 Appraisals of evaluation in remote Indigenous settings

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)

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)

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)

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)
10.5 Measuring wellbeing

The end goal of government expenditures on service delivery is to improve the wellbeing of those who receive the services, either directly or indirectly. There is no single measure of wellbeing in remote and discrete communities and limited data is available for a group of partial indicators.

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). 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).

In an Australian context, Les Malezer, from the National Congress of Australia’s First Peoples (Malezer 2012, p. 75) provides a series of measures that Aboriginal and Torres Strait Islander people may be interested in:

- political development (governance and leadership)
- property rights (ownership of land and resources)
- cultural security (exercise of responsibility over territories)
- heritage protection (language and indigenous knowledge)
- economic growth (lands and resources)
- social cohesion (positive reinforcement of identity, supportive families and communities and local decision-making).

Given the difficulties with attributing outcomes to any single service or program (discussed in the previous section), and the issues raised here about 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 10.4).

Box 10.4 Approaches to community wellbeing

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 jurisdictions, however, most measures use relatively simple metrics that are collected on a regular basis (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).
Unlike many other well-being indicators, the OECD does not provide a single index measure of wellbeing as this would require each indicator to be weighted to reflect the relative desires of the individuals residing in each jurisdiction.

**The Canadian Community Wellbeing Index (CWB)** — a measure of the standard of living and quality of life for all Canadian communities, including First Nations. The index is produced by Indigenous and Northern Affairs Canada (INAC).

A community’s CWB index score is a single number that can range from a low of zero to a high of 100. It is composed of data on income, education, housing conditions and labour force activity (Flanagan and Beauregard 2013). The CWB is also reasonably simple, allowing indices to be constructed for around 85% of First Nation communities in Canada.

**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.

A long and healthy life is measured by life expectancy at birth, knowledge is measured by years of schooling for adults aged 25 years and over and by expected years of schooling for children of school entering age, and standard of living is measured by gross national income per capita (Anand & Sen 1994).

The HDI simplifies and captures only part of what human development means, however its simplicity means that it can be easily constructed, with long time series available for many countries.

**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 (but aims to complement existing economic measures) and focuses on outcomes rather than inputs.

The index was developed with technical guidance from Michael Porter 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. The index is constructed by aggregating 53 separate indicators, where these are available (Stern et al. 2016).

While the SPI is a more comprehensive measure of wellbeing it is more complex to construct than the measures discussed above. The complexity and availability of data may also introduce a level of subjectivity that could make the index more controversial than other, more straightforward indices.

**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).

The GEM was developed collaboratively between researchers and Aboriginal residents from Alice Springs, Yarrabah and Cairns. It collects background information, such as age and gender, and asks respondents to complete an emotional empowerment scale and a number of empowerment scenarios. The GEM has been adapted to ensure cultural relevance.
10.6 Current practice

Our (incomplete) review of evaluations conducted in the context of remote and discrete Indigenous communities in Australia suggests that most evaluations are process evaluations. Those that attempt to demonstrate progress towards outcomes are rare; those that successfully demonstrate progress are even rarer. We are not aware of any evaluations that have successfully demonstrated the efficiency of service delivery in an Indigenous context.

Despite the obvious need for evaluation and the enormous amounts of public monies spent on services in Australia’s remote and discrete communities, it is a common complaint that there is little proper evaluation of these programs (DOFD 2010; Hudson 2016b; SCRGSP 2016a). Biddle (2014) argues that this goes beyond program evaluation, stating there is a general lack of information and data available to communities, service providers and decision-makers.

Although it is clear that evaluations of Indigenous programs and services are insufficient to paint a clear picture of what works and what does not, it is not obvious that this deficiency reflects a lack of evaluation effort on the ground. While we have not undertaken a comprehensive review of the evaluations completed in Australia, it appears that a significant amount of time, energy and money is spent on evaluation in Indigenous affairs. As an example, a search across 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 proportionate effort spent on evaluation in the Indigenous space is similar to the effort spent in the evaluation of social services in the general community.

The current approaches to monitoring and evaluation of Indigenous programs and services do not appear to consistently drive improved program performance. This may be an issue of implementation rather than of the evaluation framework.

10.7 Stakeholder views

Most stakeholders said that they see very little of the regular reporting from service providers. For them, this is a problem, because in their view service providers may manipulate data to suit reporting requirements. Stakeholders thought government should ask communities how they perceive service delivery.

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

Stakeholders also told us that there is little relationship between the level of risk or funding amounts and the reporting requirements associated with grant funding. 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. Council has had unqualified audits for three consecutive years, an indicator of sound financial management, and low risk within our organisation. (TSIRC sub. 12, pp. 17–18)*
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)

The requirement to source grant funding from multiple sources adds large compliance burdens to reporting. Exacerbating this, funding for services often must be sourced from two different 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)

While the difficulties inherent in evaluation may make it difficult to measure outcomes, stakeholders have consistently told us that reporting typically focuses on the wrong things.

Many program or project evaluations only look at the output level and do not assess the level of benefit to the community, organisation or region.

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)

While there are some good examples of program evaluation in Queensland, much evaluation work and information is not disclosed to stakeholders. This is perhaps best illustrated in the most recent evaluations of the Welfare Reforms (see case study, Box 10.6), but it also happens with smaller programs.

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)

Stakeholders also consistently raised the issue of funding for service delivery being insufficient to cover evaluation.
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.)

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)

This is a view shared by some academics. Cobb-Clark calls for the establishment of an independent, arms-length body to oversee all evaluation undertaken on behalf of government (like the Reserve Bank):

_Unfortunately, our current system generally produces poor-quality evaluations that in the end do not tell us very much. Often evaluations are conducted within the very agencies responsible for meeting program objectives. 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)
**Box 10.5 Case study—monitoring reporting in the Torres Strait**

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 that 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.

*Source: TSIRC sub. 12 p. 39-40.*

**Box 10.6 Evaluation of the Welfare Reforms**

The Welfare Reforms grew out of a partnership between the Cape York Institute, the Australian Government, the Queensland Government and the Cape York communities of Aurukun, Coen, Hope Vale and Mossman Gorge. The welfare reforms were built on the premise that the welfare system had created welfare passivity in remote Indigenous communities, undermining self-reliance and promoting dependence and dysfunction (Cape York Institute 2007).

Comprehensive welfare reforms were trialled in the four communities from 2008 and extended to Doomadgee in 2014. An evaluation framework, developed with the partners to the reforms, formed an essential part of the trials (FaHCSIA 2012). An initial evaluation found that, although the reforms had not been in place long enough to achieve their intended outcomes, they reflected a level of progress not evident in other reform programs in Queensland’s remote Indigenous communities, having significantly improved governance in participating communities (FaHCSIA 2012).

In 2012, the Australian Government commissioned a review of the wellbeing centres associated with the reforms. It found that the centres had achieved significant success in helping individuals through an immediate crisis, and had some success in achieving sustained positive behavioural change, although this was yet to translate into sustained, consistent and clearly observable improvements in outcomes at the community level (HOI 2014).

In the 2015–16 Budget the Queensland Government announced an additional $28.6 million over four years and $8 million a year ongoing to continue welfare reform. It also commissioned an investment review of the program to date. The review was not made publicly available, consultation was limited, and stakeholders told us they were not provided with copies of the review.
10.8 Moving forward to a new model

The current approach to evaluation and monitoring of service delivery in remote and discrete Aboriginal and Torres Strait Islander communities is not working as well as it could. A lack of effective evaluation is holding back the design and delivery of services that work, and service delivery remains largely unaccountable to the communities it should be serving.

To some extent, the deficiencies of evaluation outlined in the previous sections can be applied to service delivery more broadly. Many of the same issues that apply to remote and discrete Aboriginal and Torres Strait Islander 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.

A new approach to evaluation should support and empower communities to be active participants in developing the solutions that will work for them.

Monitoring and evaluation must support reforms to devolve responsibility and decision making

An 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, should 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 for 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.

Currently, communities have access to very little information, and reporting and compliance exercises are almost solely directed to 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 reform proposal must include much greater support for the provision of information to communities than exists today.
Transparency

As noted previously in the chapter, transparency is often identified as one of the keys to best practice in evaluation, but it is not always one of the hallmarks of monitoring and evaluation of government-funded service delivery in Queensland.

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—to be negotiated between government and communities
- periodic, public release of information on progress and expenditures in remote and discrete Aboriginal and Torres Strait Islander communities.

Practical approaches

In moving to a new approach for evaluating service delivery, the practical limitations discussed in this chapter need to be considered—these limitations mean that there is no single best way to conduct evaluation, and that there needs to be a greater focus on suites of programs rather than evaluations of single programs or services.

The Palm Island Community Council argues that:

*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)

This practical approach must support adaptive practice:

*Our research found evidence suggests organisations are more likely to engage with the evaluation process when it is presented as a learning tool to improve program delivery than when it is presented as a review or audit of their performance. This approach differs from traditional ideas of accountability, and involves moving away from simply monitoring and overseeing programs to supporting a learning and developmental approach to evaluation. 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)

This means that the emphasis of evaluation needs to shift—programs that have a robust program logic need to be given time to succeed and to learn from failure. Programs should be cut only when it can be demonstrated that they are not working and that there are obvious alternatives available that are more likely to succeed. Where there is ambiguity, benefit of the doubt should be given to continuation.
Given the small size of many communities, the focus of evaluation needs to be on the longer run. While it will remain important to track and report progress regularly, short-term changes to outcome data often provide meaningless measures of progress because of statistical noise.

The 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 evaluation reports for services it funds, either directly or indirectly.

**Comprehensive measures of progress**

Measuring progress is subjective—many of the readily available indicators of success, such as Year 12 attainment rates, may not be useful or relevant as measures of progress in remote and discrete Aboriginal and Torres Strait Islander communities. A focus on these partial measures may encourage service providers to engage in behaviours that do little to improve outcomes for remote and discrete Aboriginal and Torres Strait Islander communities—for example, Year 12 attainment rates could be improved by imposing penalties on non-attendance, but may do nothing to change the quality of education needed to improve learning outcomes.

More comprehensive indicators are required to assess wellbeing. These indicators need to be designed and measured with the involvement of the people living in remote and discrete Aboriginal and Torres Strait Islander communities—the only way of ensuring that these measures capture the things important to communities.

Importantly, a more comprehensive suite of measures is required to ensure that service delivery is providing value for money. A focus on partial indicators that do not really measure progress can mislead decision-making and cause costly program revisions.

To move to a better model, government will need to enable the collation of data that:

- are of interest to remote and discrete Aboriginal and Torres Strait Islander communities
- measure progress in a more complete way—which might include the measurement of community governance or other measures of capacity, and may need to better establish a baseline against which progress can be measured.

Rather than doing this in isolation, government needs to find ways of working with communities—this requires a delegation of decision-making and accountability to communities.

The Commission notes that 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 could 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.

**Independent oversight**

As discussed in this chapter, a common complaint is that evaluations are often undertaken, but the learnings from these evaluations are ignored, or decisions to shift or change funding are made before evaluation findings are completed (Morgan Disney 2007). If these complaints are to be overcome, any new approach needs to ensure that evaluations are de-politicised and that evaluation can be used to support and identify good service delivery practices—this will require an approach that is at arms-length from government.

Further, enabling many of the actions described above will take time. Ensuring these actions stay on track will require an independent and transparent assessment of progress.
Complementing the recommendations outlined in Chapter 7, an independent oversight body should be charged with:

- reporting on progress of an agreed reform agenda (including for evaluation)
- reporting on expenditures and outcomes in remote and discrete Aboriginal and Torres Strait Islander communities every two years.

Recommendation 8

The Queensland Government should publish expenditures made by the state in communities, including the proportion spent on indirect or ancillary functions—these should be reported every one to two years to support transparency and decision-making.

Recommendation 9

The Queensland Government should commit to an evaluation and reporting framework. This framework should support adaptive practice, facilitate accountability and empower communities by providing them with timely, useful and relevant information. To support this framework, the government should assign an independent body to:

- consult with Indigenous communities to identify the outcomes they are interested in tracking
- publish an analysis of progress of reforms and outcomes in communities every two years
- compile agency data and make this available to communities and other relevant stakeholders on a timely basis
- act as a clearinghouse for all evaluations of service delivery in communities.

The Queensland Government Statistician’s Office may collect and provide outcomes data to support this function.