Evaluating pilot programs

Contents

Introduction

Purpose

This guide is for policy and program staff and managers involved in designing and implementing pilot programs. It will help to ensure monitoring and evaluation frameworks, and related evaluation plans, are developed early and in parallel with pilot design processes.

From a monitoring and evaluation perspective, the guide suggests focus areas and key evaluation questions of specific interest for pilot programs. There are step-by-step instructions for each stage, including advice on how evaluation findings can be used while you run the pilot and after it finishes. Additional resources are also identified.

Defining pilot programs

A pilot program is a small-scale trial of a new policy, intervention or delivery approach. It is usually short-term and may apply to different demographics or geographic areas.

We run, monitor and evaluate pilot programs to:

- understand implementation

- find evidence of outcomes

- test if they are feasible

- decide whether to scale the program.

Focus areas

While similar steps apply to all evaluations, we evaluate pilot programs differently to other programs because they are usually testing a new approach.

Program evaluations typically test how well an established program was implemented and if it has achieved its intended outcomes and objectives at different stages of implementation.

Pilot programs are not established programs. They aim to test and learn about a particular approach in a particular situation. This helps to decide whether to invest more resources.

Four specific areas of interest for pilots include implementation, evidence of promise (or evidence that anticipated early outcomes have been achieved), feasibility and scalability. This section describes the evaluation questions and methods typically used to analyse these areas.

1. Understand implementation

It is important to understand how the pilot works in practice. This includes how it was delivered and what factors influenced its implementation. Without this understanding, it is hard to interpret any outcome data or make necessary adjustments. Pilots may fail due to poor implementation or unexpected barriers, rather than flaws in their conceptual design.

Key questions

You can ask questions about different aspects of implementation, such as:

- **Fidelity** – was the pilot delivered as designed? Which components matched the design? Which components varied?

- **Dosage/intensity** – did the pilot deliver the intended amount of the intervention to the participants? Did participants receive the correct amount?

- **Reach** – did the pilot reach the intended target population? Did reach vary across subgroups or geographic area?

- **Engagement** – how did participants and delivery staff engage with the intervention?

- **Experience** – how did delivery staff perceive the implementation process? How did participants experience the implementation process?

- **Barriers and facilitators** – what factors helped implementation? What factors got in the way of delivery? These factors could be resources, training, leadership support, participant characteristics or external events.

- **Contextual factors** – how did the setting influence delivery? This should include specifics such as organisational, social and political factors.

Methods

Different methods can be used to answer these questions.

For example, you can review documents, observe activities or analyse administrative data. You can also conduct interviews or focus groups with staff and participants.

2. Find evidence of promise (early outcomes)

Pilot programs should aim to collect 'evidence of promise'. These are early signs that the program, or a specific intervention or 'treatment' within it, is effective. Effective means the pilot is delivering the results or early outcomes that it was designed to achieve.

This evidence helps build confidence and get support for more development or expansion.

Key questions

You can ask questions to understand if early outcomes were achieved as expected, such as:

- Are participants' knowledge, attitudes, skills, or behaviours changing? Do these changes align with the intended outcomes?

- Do participants report getting benefits from the intervention?

- How do the pilot intervention's results compare with the results of the usual practice?

Methods

You can use the following methods, amongst others, to collect evidence of promise.

- Analysis of program output data (such as completion of training modules, use of resources) to understand what was done.

- Interviews, focus group discussions and surveys to understand what changes and benefits people noticed.

- A randomised trial or quasi-experimental approach to assess how well elements of the program work.

3. Test feasibility

Pilots test whether a proposed intervention is practical in a real-world setting. They test if the intervention works as intended in the target context. You should test:

- whether it's practical – can you deliver it effectively with available resources?

- whether people accept it – do those involved see it positively and value it?

Even if an intervention sounds good in theory, it may fail if it is too complex or costly to deliver. It may also fail if target users or implementers find it burdensome or inappropriate.

Key questions

- Were delivery partners willing to adopt or adapt the program?

- How easy or difficult was the intervention for delivery staff to deliver?

- Did delivery staff find the intervention acceptable? Did it fit with existing workload?

- Did participants find the intervention acceptable? What did they find it useful for? Did they find any burdens? Was the service appropriate?

- Did participants complete the intervention? Did they follow the program requirements? If not, why not?

- Could the program access the resources it needed? Were the resources adequate? Resources include time, funding, personnel and materials.

Methods

It is important to compare planned and actual resource use when assessing if something is practical and feasible. You can use the following methods, amongst others, to assess feasibility.

• Surveys can measure ease of use, satisfaction and burden.

• Investigate experiences and perceptions in interviews and focus groups.

• Analyse program data including costs, resource use, participation and completion rates.

4. Decide whether to scale the program

Pilot evaluations assess the potential to apply an intervention more broadly. This can include:

• expansion – reaching more people in similar settings

• adaptation – adapting the approach for different settings.

During this assessment, consider sustainability and factors necessary to maintain how well the intervention works at a different scale. Consider whether you can scale it while you monitor the pilot phase, not just at the end.

Key questions

To assess scalability, it is important to examine a range of different elements.

Intervention

• **Effectiveness** – has the pilot produced enough 'evidence of promise'?

• **Clarity and simplicity** – is the intervention well-defined and easy to understand? Is it easy to implement?

• **Adaptability** – can you adjust the pilot to fit different contexts? Will adjusting reduce effectiveness?

• **Relative advantage** – is it demonstrably better than existing practices?

The future delivery organisation

• **Capacity** – what capacity do you need to support scaling? This can be skills, leadership, management structures and resources.

• **Training needs** – what level of training do you need for scale-up? How much technical support do you need for scale-up?

• **Fit** – how well does the intervention align with the organisation? This includes the organisation's mandate, priorities and existing systems.

Context

- **Socio-cultural appropriateness**: does the intervention fit the cultural and social norms and values of potential target populations?

- **Policy and political support**: is there sufficient political will and a supportive policy and regulatory environment for scaling?

- **Stakeholder engagement and demand**: is there evidence of demand from potential beneficiaries? Are key stakeholders (community leaders, other entities) supportive?

Resources

- **Costs**: how cost effective is the intervention compared to existing or other potential solutions?

- **Financial sustainability**: is there a realistic plan for longer-term funding at scale?

- **Workforce**: is the necessary workforce available and sustainable?

Methods

To assess scalability, you can:

- analyse pilot monitoring and evaluation findings – this could include:

  – process, feasibility, and cost data

  – results from a randomised trial or quasi-experimental approach, if applicable

- do targeted stakeholder consultations

- analyse the policy and resource context

- use a scaling checklist.

Step-by-step: evaluation planning for a pilot program

Start early – build evaluation planning into pilot program design

Start planning how you will monitor and evaluate the pilot during early program design.

Embedding evaluation planning and evaluation tools in from the beginning ensures that:

- you have clear goals and measurable outcomes for your pilot

- you know what data to collect and can build it into monitoring and evaluation system

- you set aside the resources you need for monitoring and evaluation.

Follow these steps to plan how you will monitor and evaluate your pilot:

• Step 1: Define pilot objectives and scope

• Step 2: Create a theory of change and/or logic model

• Step 3: Write key evaluation questions

• Step 4: Choose the right evaluation methods and performance measures

• Step 5: Create a monitoring and evaluation plan for the pilot

• Step 6: Allocate resources for pilot monitoring and evaluation

• Step 7: Build in continuous feedback loops

• Step 8: Establish mechanisms for regular review and decision-making.

Step 1: Define pilot objectives and scope

Define pilot objectives and scope by writing clear goals and measurable outcomes (focussed on early signs or 'evidence of promise' you would expect to see if the pilot is working as intended).

These should explain what your pilot wants to test, learn, or demonstrate. For example:

• "Test if we can deliver [X services] through community health workers in [Location Y] over 12 months."

• "Find early signs that [Intervention Z] helps improve short-term results (like what participants know or how they change their behaviour)."

• "Find the main problems and changes we need for [intervention W] before we roll it out wider."

• "Test if [Intervention X] helps staff do [Policy D] better."

The scope of the pilot must also be clearly defined, specifying:

• what parts you are testing

• who you are targeting

• where you will run it

• how long it will run.

Step 2: Develop a theory of change and/or logic model

The theory of change or logic model for a pilot program explains:

• how your planned inputs (resources) and activities will create outputs (direct products or services)
• how your outputs will create short-term outcomes (early changes and evidence of promise)
• how outcomes help you reach your pilot's goals
• what you assume will make this pathway work.

Your logic model helps you write key evaluation questions and identify performance measures. It shows you where you need monitoring and evaluation data. It also highlights your key assumptions.

You can find more help on creating Theory of Change and logic models in the Commonwealth Evaluation Toolkit.

Step 3: Write key evaluation questions

Key evaluation questions help ensure that the data collected and analysed during implementation of the pilot answers the most important questions.

Example key evaluation questions on implementation process

- What were the main activities? Where did you run them?

- How well did you follow your intended design? Did different sites run it differently?

- How did people learn about and join the program?

- How much did participants take part in and complete the program activities?

- What were the main barriers and things that helped during delivery?

- Was the planned amount of intervention right?

- What did the pilot program cost?

- What factors in the setting affected how you ran the pilot? How?

Example key evaluation questions on results and how well it worked

- Have participants' knowledge, attitudes, skills, or behaviours changed? Do these changes match what your pilot's Theory of Change describes?

- What benefits do participants and staff think the pilot gave?

- What proof do you have that the pilot, or parts of it, work well?

Example key evaluation questions on whether you can scale it up

- Can you scale up or replicate the pilot model as it is now?

- What changes would be needed to make it work on a larger scale or in different places?

- What did the pilot cost, and what would it cost to scale up?

- What factors in institutions, policy, or resources might help or stop future scaling?

Further guidance on developing key evaluation questions can be found in the Commonwealth Evaluation Toolkit.

Step 4: Choose the right evaluation methods and performance measures

Choose your evaluation approaches and methods based on your key evaluation questions. Be realistic about your pilot's timeline, budget, and available skills. Many of the possible methods were detailed in the focus areas for monitoring and evaluating pilots.

Different evaluation questions require different evaluation methods. Consider using mixed methods. This means you bring together quantitative and qualitative data and analysis.

For example, you could use impact evaluation along with descriptive insights from interviews and focus groups. This helps you understand your quantitative information.

Developmental evaluation may also work well in some contexts because it focuses on identifying helpful learnings and implementing quick changes during delivery.

Table 1 Overview of key evaluation questions, evaluation types and methods

Focus area	Indicative Key Evaluation Questions	Evaluation type	Methods
Understanding implementation	• How well did you follow your intended design? • How much did participants take part in and complete the pilot activities? • Was the planned amount of intervention right? • What contextual factors influenced the implementation of the pilot? How?	Process evaluation	• Observation • Interviews and focus groups • Surveys • Analysis of program data • Participatory research
Evidence of promise (early outcomes)	• Were the intended outcomes of the pilot program achieved? • Are there changes in participants' knowledge, attitudes, skills, or behaviours that match the intended outcomes of the pilot?	Outcome evaluation	• Interviews and focus groups • Surveys • Analysis of program data • Analysis of administrative data • Participatory research
Evidence of promise (early outcomes)	• Can you show that the pilot, or elements of it, achieved the intended results (show it caused them)? • How do the outcomes compare to business-as-usual approaches?	Impact evaluation	• Experimental Designs • Randomised trials • Quasi-Experimental Designs • Regression discontinuity design • Difference-in-differences

Focus area	Indicative Key Evaluation Questions	Evaluation type	Methods
Feasibility	• Were delivery partners willing to adopt and/or adapt the pilot program? • How acceptable was the intervention to participants and delivery staff? • Did participants complete intervention? • Were the necessary resources (time, funding, personnel, materials) available and adequate?	Process evaluation	• Process evaluation • Observation • Interviews and focus groups • Surveys • Analysis of program data • Participatory research
Scalability	• Has the program been cost-effective (compared to alternatives)? • Do the benefits of the program outweigh the costs?	Economic evaluation	• Economic evaluation • Cost effectiveness analysis • Cost benefit analysis • Multi-criteria analysis
Scalability	• Is there sufficient 'evidence of promise' from the pilot? • Is the intervention better than existing practices at achieving desired outcomes? • Is the intervention well-defined and feasible to implement? • Does the intervention fit the cultural norms and values of the potential target populations? • Is there evidence of demand? • Are key stakeholders supportive? • How easily can the intervention be modified to fit different contexts? • What resources are required to support scaling? • How well does the intervention align with the implementing organisation's mandate, priorities, and existing systems? • Is there enough political will and a supportive policy environment for scaling?	Elements of all the above types of evaluation	• Analysis of pilot monitoring and evaluation findings • Conduct targeted stakeholder consultations (policymakers, potential implementers, funders) • Analyse the policy and resource context

Further guidance on impact evaluation and developmental evaluation can be found in the Resources section of this document.

Establish robust data collection early

Indicators that will help to answer key evaluation questions should be determined and data collected on them from the outset of the pilot.

Indicators will depend on the key evaluation questions but are likely to include indicators of:

- **Inputs and activities:** relating to resources used and activities completed

- **Outputs:** the products or services delivered, how many people your pilot program reached (like number and type of participants)

- **Process quality:** how well you followed your plan, what staff think, how participants took part and how satisfied they were

- **Feasibility**: how easy it was to deliver, how you used resources (planned versus actual), how many people stuck with it

- **Early/short-term outcomes:** changes in knowledge, attitudes, skills and behaviours for the participants

- **Scalability factors:** how acceptable the intervention was, how satisfied people were with results, stakeholder support, cost data.

A monitoring and evaluation matrix can help you identify and define measures that link to your Key Evaluation Questions. It will also show the related data sources and who is responsible for collection. You can find a template for a 'data/evaluation matrix' in the Commonwealth Evaluation Toolkit [here](#).

Step 5: Create a monitoring and evaluation plan for the pilot

The monitoring and evaluation plan is a working document that brings together the elements listed above. It should include:

- **Overview**: summary description of the pilot program, including objectives, scope and desired outcomes

- **Logic model/theory of change**

- **Key evaluation questions**

- **Monitoring and evaluation matrix**:

  – indicators (linked to key evaluation questions)

  – indicator definitions

  – baseline data & source

  – targets/benchmarks (if realistic for pilot)

  – data collection methods and tools

  – how often you collect data

  – who is responsible for collection.

- **Data analysis plan** (describing how you will analyse data for each measure and key evaluation question)

- **Reporting plan**: reports/outputs, target audiences, format, frequency

- **Timeline**: key monitoring and evaluation activities mapped against the pilot timeline

- **Budget**: estimated costs for monitoring and evaluation activities

- **Roles and responsibilities**: who is responsible for monitoring and evaluation oversight, data collection, analysis, reporting

- **Ethics considerations**: plans for informed consent, data privacy, confidentiality, ethics review (if needed).

Pilots often involve ongoing learning and adjustment. You should view your monitoring and evaluation plan as a living document that you can adapt to changing circumstances, decisions or emerging new questions.

Step 6: Allocate resources for pilot monitoring and evaluation

You need adequate resources to do meaningful monitoring and evaluation. Pilots will likely need a higher proportion of program budget than you typically set aside for program monitoring and evaluation. This reflects the focus on intensive monitoring and learning for a new policy, intervention or delivery approach. It also reflects how important it is to create strong evidence for scaling decisions and the likely need to use a range of methods.

Your resourcing decisions will depend on your context. Resourcing should be proportional to your pilot's scale, complexity, risk and opportunities.

Resourcing should be considered as part of your initial pilot proposal and planning process. As well as financial resources, consider:

- **Staff time**: dedicated time for data collection, analysis, reflection and coordination

- **Expertise**: access to appropriate monitoring and evaluation skills. This might involve internal staff with monitoring and evaluation responsibilities, dedicated evaluation units, or external consultants or evaluators. If the evaluation type or data collection method requires specialist skills, such as impact evaluation, involving the right people early will help to ensure evaluation findings are robust and credible.

- **Data systems and tools** for data collection (like survey platforms, data agreements), storage, and analysis.

Step 7: Build in continuous feedback loops

You get the full value of pilot monitoring and evaluation when you use findings to inform learning, adaptation, and strategic decisions about the intervention's future and scale. You should establish clear processes for feedback, interpretation, and linking evidence to action.

Use evaluation evidence for adaptive management and decision-making

Unlike evaluations that happen at the end of a project cycle, monitoring and evaluation of a pilot should include feedback loops that provide rapid feedback to inform decision-making during delivery. This allows for adaptive management. You can then make timely adjustments to the pilot based on emerging data and insights.

Ways to get continuous feedback include:

- **Real-time data tools**: using dashboards or simple tracking systems, where possible, to give immediate visibility on key metrics

- **Regular review of monitoring data**: reviewing key process measures (like reach, how well you followed your plan, participant feedback) at frequent intervals

- **Interim reports and briefings**: sharing early findings and observations from monitoring and evaluation activities with the implementation team and managers

- **Regular reflection meetings**: involving implementers, managers, and monitoring and evaluation staff in regular reflection meetings creates opportunities to discuss monitoring data, share experiences, identify challenges, and solve problems together.

Approaches like developmental evaluation and real time or rapid evaluation clearly build in repeated cycles of data collection, reflection, and adaptation.

Step 8: Establish mechanisms for regular review and decision-making

Interpret findings collaboratively to improve pilot program

You should analyse monitoring and evaluation findings to pinpoint areas where you can improve the pilot intervention or how you deliver it. This might involve actions like:

- refining program content

- adjusting how you deliver it

- improving training or support materials

- addressing delivery bottlenecks.

You must involve key stakeholders, especially front-line implementers and participants, to interpret findings. Their insights can help explain data and identify practical solutions, and make sure that proposed improvements are appropriate for the context and feasible.

Inform decisions on scaling, adaptation, or discontinuation

At the end of a pilot (or at key interim points), you need to bring together the monitoring and evaluation evidence to inform strategic decisions about the pilot program's future.

Potential decisions could include:

- **Scale-up**: where the evidence suggests the intervention is practical, acceptable, shows promise, and key scaling factors appear favourable (perhaps with minor adaptations). The monitoring and evaluation findings would inform the plan for wider implementation.

- **Adaptation**: where the pilot showed potential but also revealed flaws in the model or delivery approach that you need to address before scaling. Monitoring and evaluation findings would guide the necessary modifications. This may lead to a subsequent, refined pilot phase.

- **Further testing**: where the pilot showed it was practical and promising, meeting readiness criteria for more rigorous evaluation to establish causal impact before large-scale investment (for example, through a randomised trial or quasi-experimental approach).

- **Stopping**: where the pilot revealed fundamental problems with practicality, lack of acceptability, absence of any evidence of promise, insurmountable barriers to scaling, or significant negative unintended consequences.

Regardless of the decision, you should clearly document the process and the evidence behind it to ensure transparency and accountability.

Communicate findings effectively

How useful monitoring and evaluation findings are depends on effective communication with the relevant audiences.

You should tailor communication products to meet the needs of different audiences. For example, senior decision-makers may need brief summaries focusing on key findings and recommendations for strategic choices, while implementation teams may need more detailed process findings to guide operational adjustments.

You may need a range of communication formats including:

- comprehensive reports

- focused briefs

- presentations and workshops.

Conclusion

Rigorous and proportional monitoring and evaluation approaches are critical to maximise the learning you can get from pilot programs. It can help:

- reduce the risk of investing in unproven interventions

- support improved program design through step-by-step refinement

- build a stronger evidence base for justifying future actions, whether that be scaling up, adapting, or stopping an initiative.

Key features of an effective pilot program evaluation include:

- plan monitoring and evaluation early and integrate it into pilot program design

- use mixed methods appropriate to your pilot's context and learning objectives

- communicate and use findings in a timely way to support adaptation and strategic decision-making.

Ultimately, pilot programs help policymakers test new approaches to improve outcomes for program participants. Strengthening pilot monitoring and evaluation contributes to more effective, evidence-informed programs and enhances the value delivered through public resources.

Resources

Australian Centre for Evaluation, Economic evaluation

Australian Centre for Evaluation, Impact evaluation

Australian Institute of Family Studies, Developmental Evaluation

Better Evaluation, Specify the key evaluation questions - Rainbow Framework