ROADMAP TO SOCIAL IMPACT

YOUR STEP-BY-STEP GUIDE TO PLANNING, MEASURING AND COMMUNICATING SOCIAL IMPACT

SECOND EDITION

Prepared by the Centre for Social Impact

Cat Street, Lisette Kaleveld, Paul Flatau, Rhiannon Parker, Andrew Joyce, Nada Ibrahim, Arminé Nalbandian

CENTRE *for* SOCIAL IMPACT

UNSW SYDNEY

THE UNIVERSITY OF WESTERN AUSTRALIA

Flinders University

SWINBURNE UNIVERSITY OF TECHNOLOGY

Acknowledgement of Country

We respectfully acknowledge the Traditional Owners of the land on which The Centre for Social Impact is located in Perth, Sydney, Melbourne and Adelaide. We pay our deep respect to elders past and present and extend that respect to all Aboriginal Torres Strait Islander peoples.

Address for Correspondence

Professor Paul Flatau
Centre for Social Impact
The University of Western Australia
paul.flatau@uwa.edu.au

Suggested citation: Street, C., Kaleveld, L., Flatau, P., Parker, R., Joyce, A., Ibrahim, N., Nalbandian, A.. (2025). *Roadmap to outcomes measurement. Your step-by-step guide to planning, measuring and communicating social impact, second edition*. Centre for Social Impact.

This publication is a second edition of the Roadmap.

Ramia, I., Powell, A., Stratton, K., Stokes, C., Meltzer, A. & Muir, K. (2021). *Roadmap to outcomes measurement. Your step-by-step guide to planning, measuring and communicating social impact*. Centre for Social Impact.

Keywords: social impact, outcomes measurement evaluation.
Publisher: Centre for Social Impact.
Format: PDF.
URL: https://www.csi.edu.au/tools-and-guides/roadmap-to-social-impact

Centre for Social Impact

The Centre for Social Impact ignites positive, lasting social change for people and communities through research education and engagement. It is a national collaboration of four leading Australian universities, working together to reduce social inequities and create a future where everyone can thrive.

We gratefully acknowledge the support of the Professor Peter Shergold, Social Impact Fund, and extend our thanks to the AMP Foundation as its major contributor.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

PURPOSE AND CONTEXT

PURPOSE

The *Roadmap to Social Impact: Your step-by-step guide to planning, measuring and communicating social impact* ('the *Roadmap*') is a guide to support you and your organisation through the process of outcomes measurement and evaluation for the programs you run. This guide helps you demonstrate the social impact your program achieves.

For simplicity, we use the term 'program' throughout the *Roadmap*. However, your work may involve implementing a suite of programs, a policy, strategy or systems change effort. Regardless of the context, if your work is aimed at achieving social impact and you seek to understand the outcomes of your actions, this guide is for you.

The *Roadmap* outlines the key components involved in a typical outcomes measurement and evaluation journey within the social impact sector. Each chapter covers key social impact concepts, their importance to outcomes measurement and evaluation, and how they are applied in practice. However, this is not an exhaustive 'how to' guide detailing every approach and method. We recognise that there are numerous ways to do outcomes measurement and evaluation, and that many resources provide more detailed information on specific approaches and methods.

The *Roadmap* serves as an entry point, offering:

- An introduction to the vocabulary of outcomes measurement and evaluation.
- An overview of key approaches, methods and tools.
- Examples and case studies.
- Key considerations for each stage of your outcomes measurement and evaluation journey.

It is our hope that the *Roadmap* will equip you and your organisation with the necessary knowledge and tools to confidently progress your outcomes measurement and evaluation journey. Links to supplementary resources are provided throughout the document and in the appendices.

CONTEXT

Who this guide is for

The *Roadmap* is intended for anyone working to create positive social impact in Australia and who seeks to understand their contribution at the individual, organisational, community and societal levels. 'You' (the reader) may be implementing a program, or be a program manager, a leader, or any stakeholder involved in outcomes measurement and evaluation. You may also simply be interested in learning more about how to measure and evaluate social impact.

History of the Roadmap

The second edition of the Roadmap represents the Centre for Social Impact's (CSI) most up-to-date resource on social impact outcomes measurement and evaluation. It replaces the previous version of the Roadmap published in 2021[1]. CSI's The Compass: Your guide to social impact measurement[2] ('the Compass') was an earlier CSI resource that helped practitioners understand how to conduct social impact outcomes measurement and evaluation. This updated Roadmap incorporates content from the previous Roadmap and the Compass while reflecting the latest way of thinking, as the practices of outcomes measurement and evaluation develop, mature and respond to contemporary contexts.

INTRODUCTION

Outcomes measurement and evaluation aims to measure, assess and understand the value of social impact initiatives.

Increasingly, there has been a shift away from seeing outcomes measurement and evaluation activities as a compliance-driven obligation to a core activity that supports learning, innovation, strategic decision making and advocacy. Outcomes measurement and evaluation are important because they:

- Demonstrate value by providing evidence of a program's effectiveness.
- Motivate teams by showing the tangible difference their work makes.
- Support decision-making for funding and resource allocation.
- Link activities to an organisation's vision and values.
- Enhance communication with stakeholders, funders and the wider community.

The *Roadmap* will guide you through a **nine-step journey** to outcomes measurement and evaluation, using a structured approach, with each step focusing on a key component of the outcomes measurement and evaluation process.

THE NINE STEPS OF THE ROADMAP

1 **Establish purpose**
- Define the problem your program addresses.
- Align activities with your vision, mission, goals, and objectives.
- Clarify your reasons for wanting to measure outcomes and evaluate.
- Determine how measurement links to other activities, events, funding cycles or decision-making.

2 **Plan**
- Foster a culture of learning.
- Identify stakeholders and their roles.
- Consider and plan for ethical requirements.
- Secure resources for outcomes measurement and evaluation.

3 **Clarify program design**
- Develop a theory of change.
- Build a logic model (inputs, activities, outputs, and outcomes).

4 **Understand what to measure and evaluate**
- Prioritise and develop evaluation questions.
- Select an appropriate evaluation type.

5 **Develop an outcomes measurement and evaluation framework**
- Identify priority outcomes for measurement.
- Identify relevant indicators.
- Identify data sources to quantify and qualify indicators.

6 **Collect data**
- Select quantitative and/or qualitative methods.
- Follow a planned data collection process.
- Address the ethics and politics of data collection in outcomes measurement and evaluation.

7 **Analyse data**
- Apply appropriate analytic frameworks.
- Assess change and impact.
- Recognise bias and ensure data quality.

8 **Undertake knowledge translation**
- Communicate results effectively.
- Engage stakeholders with insights.
- Use findings to enhance social impact.

9 **Embed evaluation and evaluative thinking**
- Build outcomes measurement and evaluation skills and capabilities.
- Facilitate evaluative thinking.

THE LANGUAGE OF SOCIAL IMPACT

The language of social impact can be confusing at times. The world of outcomes measurement and evaluation and related definitions and concepts have evolved out of diverse disciplines, meaning the language is not always standardised and different definitions exist. Still, understanding terminology is an important starting point. The box below provides definitions used in the *Roadmap* for some of the terms you might have come across.

In the *Roadmap* we distinguish between the terms 'outcomes measurement' and 'evaluation' which are presented as complementary but distinct processes:

- *Outcomes measurement* refers to a systematic way to identify and track changes resulting from a program.
- *Evaluation* takes outcomes measurement one step further. It assesses the **merit, worth, and significance** of those changes. Evaluation seeks to answers critical questions such as:
  - Did the program achieve it's intended impact?
  - What was the impact of a program against what would have happened otherwise?
  - Would these changes have occurred without the program?
  - Was the program delivered effectively and efficiently?

To be able to answer questions such as these, a robust design is needed that identifies the achievement of outcomes of the program in question, but also examines what might have occurred had the program not been implemented (what is referred to as the counterfactual). Evaluation also explores other elements of program merit, worth and significance beyond outcomes such as design, implementation, efficiency, cost-effectiveness and value. Different types of evaluation serve different purposes as shown in Table 1 (see next page).

This guide will unpack these concepts in more detail and provide practical insights into how they fit into the broader outcomes measurement and evaluation process. A glossary at the end of this document offers additional definitions and clarifications.

The language of social impact

- **Outcomes measurement:** A systematic way to describe and measure the outcomes of a program.
- **Evaluation:** A systematic process of assessing and understanding the merit, worth or significance of a program, by combining evidence and values[3]. Evaluations address questions relating to different aspects of a program such as:
  - Whether it correctly diagnoses the problem it intends to address.
  - Whether the design of the program is suitable to address the problem.
  - Whether implementation occurs in a manner consistent with the design.
  - How a program was implemented, what effects it had, for whom, how and why.
  - Whether it improves outcomes for the target population.
  - Whether it does so at an acceptable value and cost to society.[4]
- **Outcomes evaluation:** The assessment of the changes or outcomes resulting from the implementation of a program. It includes both intended and unintended outcomes for a range of stakeholders engaging in a program.
- **Process evaluation:** The investigation of the extent to which a program was implemented as planned. It helps understand why changes occurred.
- **Social impact:** The intended and unintended social consequences, positive and negative, of programs and any social change processes invoked by these.[5]
- **Impact evaluation:** The assessment of the extent to which long-term, sustained changes resulted from a program. The term 'impact evaluation' is also often used as an umbrella term to cover all levels of outcomes whether over the short- and medium-term or the long-term.

Table 1:
Different types of evaluation and their purposes.

Type of evaluation	Purpose
Outcomes evaluation	Assesses intended and unintended program effects.
Process evaluation	Examines program implementation and delivery.
Economic evaluation	Analyses cost-effectiveness, cost-benefit, and value.
Impact evaluation	Measures long-term, sustained changes from a program but often covers all levels of outcomes including short-term, medium-term and long-term.

PRACTICE YOUR SKILLS

To help translate theory into practice, the *Roadmap* includes **case studies** that illustrate real-world applications of outcomes measurement and evaluation.

THE '*SPORT*' EXAMPLE

Throughout this guide, we will use a **fictional case study**, *Sport*, to demonstrate how key concepts apply in practice.

- *Sport* is a program providing **free after-school sports activities** for primary school children in low-socioeconomic areas.
- The program's goal is to **increase physical activity** by offering children the opportunity to exercise two to three times a week, participating in both team and individual sports on school grounds.

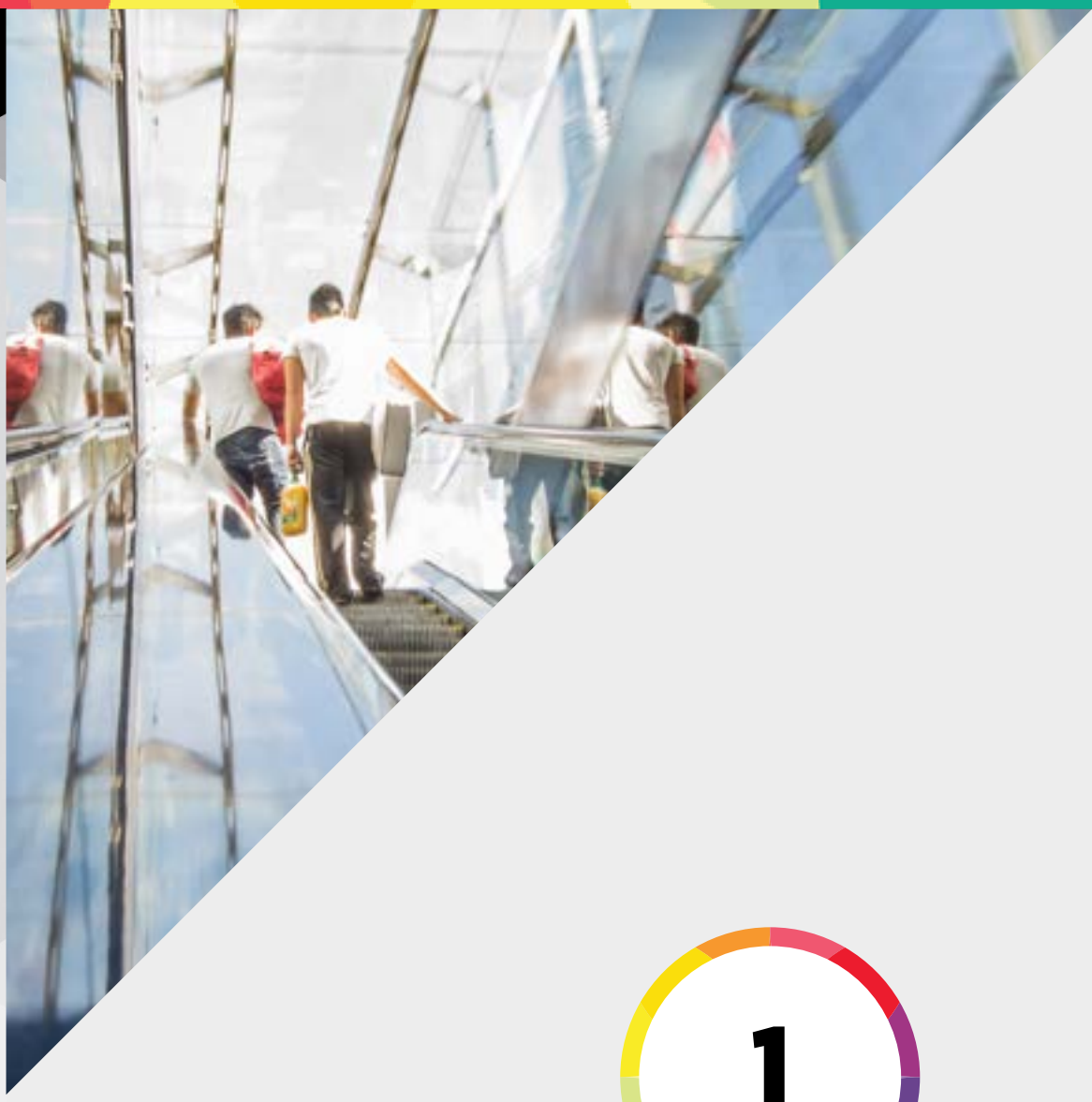
In addition, to support behaviour change the program delivers:

- **Information sessions** for students, teachers and parents on the benefits of an active lifestyle.
- **Printed resources** covering topics such as healthy sleep habits, nutritious eating and outdoors activity.

The program is free of charge for schools and students, funded by the local government, and delivered by a not-for-profit in partnership with the Department of Education and Department of Health.

Each chapter will reference *Sport* to illustrate how outcomes measurement and evaluation principles can be applied to real-world scenarios.

Keep an eye out for this symbol — it identifies sections in the report relating to *Sport*.



1

ESTABLISH PURPOSE

WHY PURPOSE MATTERS

Outcomes measurement and evaluation are most effective when planned from the outset rather than introduced as an afterthought. Ideally, planning for outcomes measurement and evaluation begins at the program design stage, ensuring clarity around what needs to be measured and why.

To build a strong foundation for measurement and evaluation, you need to be clear about:

- What *problem* your program is trying to solve.
- What your program's *intended outcomes* are.
- The *purpose* of measuring and evaluating these outcomes[6].

Without a well-defined purpose, it becomes difficult to determine what should be measured, how success should be assessed, and how insights should be used.

UNDERSTAND YOUR PROBLEM AND THE SYSTEM IT OPERATES IN

What problem are you trying to solve?

Social problems are often complex and multifaceted, or even wicked[7], with multiple causes and interrelated effects. They often cannot be resolved without multiple, simultaneous and collaborative efforts.

Problem analysis is a process that helps to:

- Understand the entrenched nature of social issues.
- Identify the 'root causes' or system drivers.
- Map potential solutions.

It can also help you identify potential partners or complementary programs that (should) work alongside your program to address the broader problem. Problem analysis also will also shed light on the underlying systems-level determinants and how your particular focus of action fits into the broader scheme of things.

Both causes and effects should be evidence-based, meaning they should be based on existing research and literature and other credible sources of knowledge and expertise including the lenses of knowledge provided by practitioners and lived experience.

Equity considerations in problem analysis

Problem analysis should always think about equity, that is, the historical, structural and contextual factors that affect people targeted by a program. This includes:

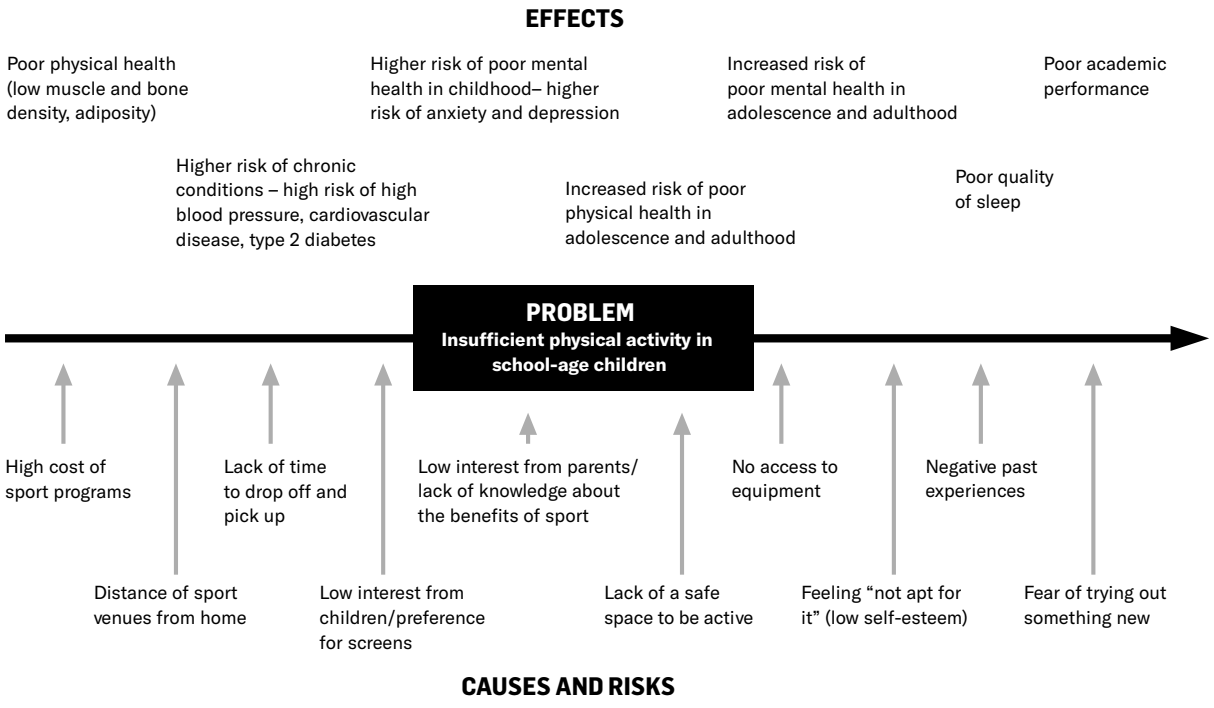
- Assessing who benefits most and least from different program designs.
- Ensuring the needs of marginalised communities are centered in accountability, learning and decision-making.
- Understanding how outcomes measurement and evaluation can contribute to a more equitable society.

Case study: Problem analysis for the fictional case study, Sport program

We now look at *Sport* to illustrate problem analysis. The core issue the program addresses is insufficient physical activity among school-aged children. This issue is part of a broader problem—an unhealthy lifestyle— which is influenced by factors such as poor diet, inadequate sleep, extended screen time and structural or policy level issues such as possible high cost of sport participation, parental support, access, infrastructure and cultural factors affecting lifestyle behaviours. While insufficient physical activity among school-aged children is one part of a larger problem, *Sport* is specifically designed to tackle low levels of physical activity, rather than addressing all factors that lead to an unhealthy lifestyle.

The problem analysis diagram (see Figure 1 below) helps identify and visualise the root causes of the problem *Sport* seeks to address, as well as the effects the program aims to alleviate or eliminate. Your program or organisation may focus on addressing some causes and alleviating some effects of the problem you identified. Taking a holistic approach to problem analysis helps you understand the context in which your program operates, clarify the specific aspect of the problem you are addressing, and identify the potential outcomes you expect to achieve (alleviation of which effects). It may also highlight the need for collaboration or partnerships to help you address challenges that cannot be resolved by your program alone and therefore lead you to a more systems-based approach.

Figure 1:
Problem Analysis Diagram – fictional case study, *Sport* program



Sources: Hu et al. (2021)[8], Clearinghouse for Sport (2024)[9].

Note: This problem tree is for illustrative purposes and may not include all causes and effects related to physical activity in school aged children. Similarly, not all causes and effects may apply to all school-aged children.

Systems thinking in outcomes measurement and evaluation

A system is an interconnected set of elements that is coherently organised in a way that produces predictable results.[10] Systems thinking is a holistic method for understanding positive and negative influences on a problem and identifying the 'big levers' for creating change. It identifies problem influences at the individual, household, community, infrastructural, political and societal level[11] and the stakeholders behind these influences. When one element of a system changes, the other parts may be affected and, in the end, the stability of the whole system.
Systems thinkers:[12]

- Seek to understand the big picture.
- See patterns in the system.
- Recognise how a system's structure causes its behavior.
- Identify cause and effect relationships.
- Surface and test assumptions.
- Find where unintended consequences might arise.
- Find leverage points to change the system.
- Resist making quick conclusions.

Systems thinking has been found useful for evaluation of programs, such as the *Sport* program, where there are multiple causes contributing to the issue, and a diverse range of solutions needed (problems such as these are often referred to as complex or wicked problems).[13]

In evaluating programs that address complex social issues, linear approaches and fixed indicators may not be suitable. What is needed instead is ongoing learning and adaptation, and privileging local knowledge, diverse viewpoints and the perspectives of those affected with lived experience of the issue.[14,15]

When evaluating programs that address complex social issues, linear approaches and fixed indicators may not be suitable. As shown in our *Sport* example in Figure 2 (see next page), applying systems thinking supports learning about the complex causes of social problems, the effects of programs and their longer term impacts within systems and how innovation and systems change can be implemented. For example, while Sport is targeted towards behaviour change in school-aged children, there may be opportunities to address other causes that relate to the wider system, for example subsidizing the cost of sports programs in the broader community and promoting active participation rather than competition in schools.

What is the big picture or the system?

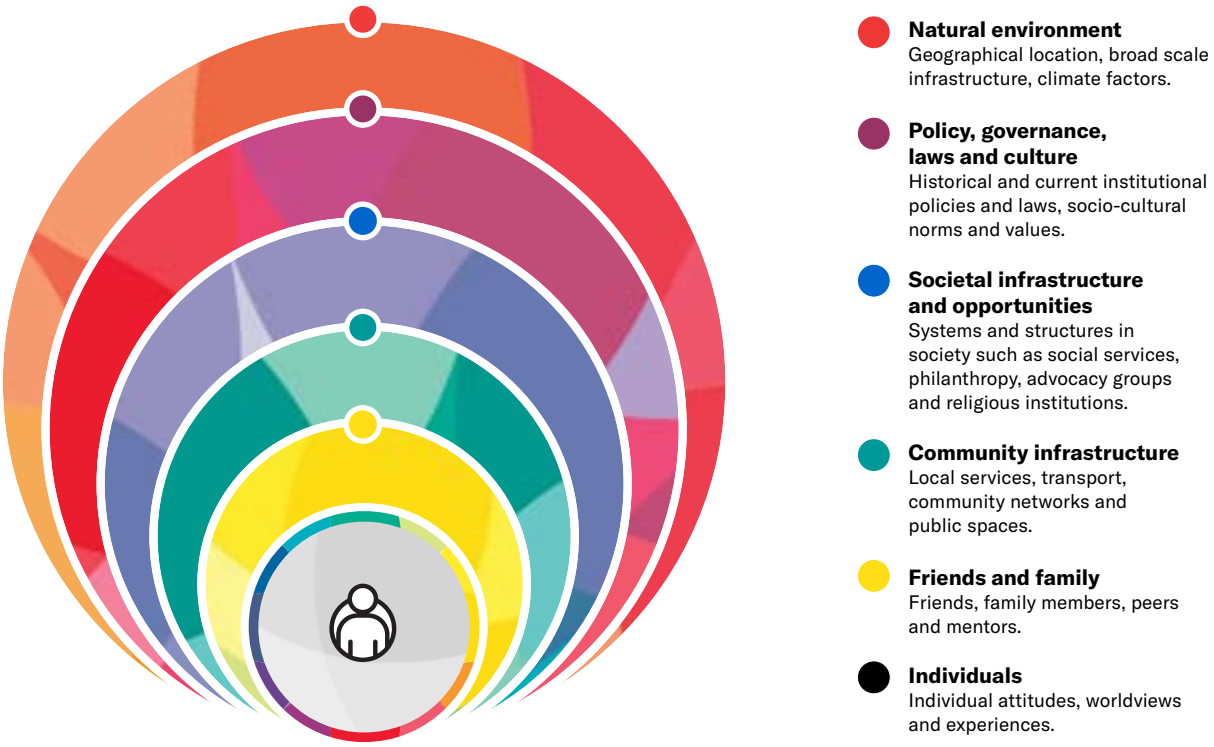
Social problems are not isolated, they exist within systems. You should think about the wider system in which a problem exists. You will need to consider the various groups or stakeholders[i] who exist in the system and how they relate to the problem, your program and to each other. Figure 2 below (adapted from Dass et al. (2020)[16]) will help you visualise potential stakeholders or elements of the system you or your beneficiaries may be engaging with, the nature of the interaction and your levers for change.

You should think about ways to incorporate the perspectives of the beneficiaries and understand how the various layers of the system affect them[17].

For example, some elements of the system for our case study are: the student, their family and home environment, the school, services available, accessibility and past experiences. These elements interact and reinforce each other while representing causes and effects of the problem.

You should also consider in your system the elements that interact with your program, such as supporting partners, other agencies or various groups you interact with such as direct and indirect clients, funders, volunteers. All these can inform the causes and effects of the problem and help you identify how your program can contribute to resolving the problem.

Figure 2:
System thinking framework applied to the fictional case study, *Sport* program



i Stakeholder analysis is further discussed in the next section, from the perspective of stakeholders engaging in outcomes measurement and evaluation.

Case study: Systems-wide evaluation of homelessness and rough sleeping

An evaluation commissioned by the UK Ministry of Housing, Communities & Local Government aimed to identify opportunities to improve the way that the homelessness and rough sleeping system works. The evaluation used systems thinking to understand how homelessness and rough sleeping is influenced by the wider system.[18]

Interviews, systems mapping and a review of major homelessness and rough sleeping policies and programs was conducted. Fieldwork was also conducted in five local authorities that consisted of interviews and focus groups with local representatives and delivery partners, as well as individuals with lived experience of homelessness and rough sleeping. A detailed systems map of the homelessness and rough sleeping systems was developed through using existing maps, conducting workshops and refining the systems map based on feedback. The map showed the connections between different parts of the system in order to understand where interventions can have the greatest impact.

The findings showed that the system is not working as intended and local authorities are facing increasing demand from people who are in crisis. It highlighted that disjointedness in funding schemes are leading to system inefficiencies and that a greater focus on prevention of homelessness is needed. The report recommends a whole-of government approach involving agreement on a clear set of shared outcomes is needed to tackle homelessness effectively.

WHAT ARE YOUR VISION, MISSION, GOALS AND OBJECTIVES?

It is important to align your program and related outcomes measurement and evaluation work to the strategic vision, mission and goals of your organisation.[19] The **vision** is an organisation's statement of its overall ideal and the ultimate goal of its operation.[20] It describes what the future should look like. The **mission** describes 'the business' of the organisation[21] or that of a program, and is more action-oriented than the vision. It describes how that future will be achieved and while it can be formulated at both organisation- and program-level, it is often articulated at program-level as an organisation would seek to achieve its vision through several programs or initiatives. The vision will provide strategic direction and facilitate decision-making, while the mission will ensure your activities align with the overall mission of the organisation. In a nutshell, vision and mission answer the following questions:

- **What** do you seek to achieve? What is your 'perfect world'? (Vision)
- **How** will you achieve that? (Mission)

The **goals** are longer-term aspirations your organisation has for the future and indicate where your organisation's efforts are directed. Your program's objectives are more tangible, specific and measurable aspirations. Your vision, mission, goals and objectives should be aligned to the problem you are looking to resolve. The box to the right presents the vision, mission, goals and objectives of *Sport*, a program delivering a single service. It may be that your program is delivered by several organisations in collaboration. Such collaborative initiatives must have missions that align with the vision of the individual organisations that deliver them[22].

Sport vision, mission, values and goals

Vision: Healthy children, healthy adolescents, healthy adults.

Mission: Provide children with opportunities to be active two to three times per week.

Goals: Reduce lifestyle induced illnesses in children.

Objectives: Familiarise parents, teachers and students with healthy habits; increase student, teacher and parent awareness of benefits of sport; instill an active lifestyle; engage students in after school sport activities two to three times per week.

LINKING OUTCOMES MEASUREMENT AND EVALUATION WITH ORGANISATIONAL VISION, MISSION, GOALS AND OBJECTIVES

Outcomes measurement and evaluation can be used to:

- Understand how well an organization is fulfilling its vision.
- Understand alignment of a programs' goals and objectives to the organisation's vision and mission.
- Understanding the extent to which a program is achieving its mission, goals and objectives.

WHAT IS THE PURPOSE OF OUTCOMES MEASUREMENT AND EVALUATION?

All outcomes measurement and evaluation work should be guided by purpose. Purpose is the reason for doing something and provides clarity, direction and motivation. Your purpose drives your outcomes measurement and evaluation process. **Without being clear on purpose, it is hard to know what needs to be measured or evaluated and what is the best way to do this.**

Consider whether your purpose for doing outcomes measurement and evaluation is to:[ii]

- Clarify or refine program scope and implementation.
- Understand whether program objectives were achieved.
- Understand whether program outcomes were achieved.
- Provide accountability to funders and other stakeholders.
- Provide accountability to the beneficiaries of a program or to the public.
- Generate insights that other programs can use.

Outcomes measurement and evaluation provides:

- An informed judgement about the value, worth or significance of a program.
- An evidence base on different components of a program (e.g. appropriateness, effectiveness, efficiency).
- A critical tool for accountability and resource allocation decisions.
- The basis for learning and responsible policy development within organisations.
- The key ingredient for strategic learning, planning and good governance.
- Learning to enhance staff engagement and motivation.
- Data required by, and to attract, funders.

WHO IS YOUR AUDIENCE?

Before you begin, you should also **think about who your audience is**. This relates closely to your purpose and will specifically guide any deliverables that you need to include in your plans.

For example, for *Sport*, is your audience the program funders, schools or parents and families? Understanding your audience will help you to know what information you want to capture and how you can plan to make the best use of this information.

ii Note that this document outlines a general process for outcomes measurement and evaluation for social impact work. It is recognised that outcomes measurement may not always be within the scope of an evaluation.

OUTCOMES MEASUREMENT AND EVALUATION SHOULD BE CAREFULLY PLANNED

Outcomes measurement and evaluation are essential activities for all organisation, but care must be taken not to rush into outcomes measurement and evaluation without good planning. Poorly planned data collection may, for example, interfere with how participants engage in the program).

Evaluability assessment is 'the extent to which an activity or project can be evaluated in a reliable and credible fashion'.[23],[24]

Evaluability assessment tests:

- Whether it is an appropriate time to measure outcomes and evaluate.
- Whether a program is ready for outcomes measurement and evaluation.
- When outcomes measurement and evaluation would help improve the program.

At this point it is also a good idea to think about who will be responsible for doing or participating in outcomes measurement and evaluation and whether capability and capacity exists. The evaluator can be internal, or external 'in-house' (e.g. a manager or researcher at the organisation delivering the program) or external. There are advantages and disadvantages to having an internal or external evaluator[25] (refer to Section 2).

Evaluators can give recommendations on when outcomes measurement and evaluation are achievable, the tools necessary, or if evaluation is possible at all. Your organisation needs to consider outcomes measurement and evaluation from the beginning, so that, for example, data collection methods can be built into the overall program implementation plan. Your program may not be ready to be evaluated but having an outcomes measurement and evaluation plan will ensure it is achievable down the track.

Evaluability assessment involves a six-step process

- Involve key stakeholders (e.g. policymakers, managers, staff, beneficiaries) - to ensure the program theory conforms with their expectations.
- Clarify program design - ensure the relationship between inputs, activities, outputs and outcomes is as expected from the points of view of key stakeholders.
- Clarify program reality - whether the program was/is implemented according to the program design.
- Assess the likelihood that the program activities will lead to the intended outputs and outcomes.
- Agree on required changes to the program design.
- Agree about the intended use and value of future evaluation activity.[26]

OUTCOMES MEASUREMENT AND EVALUATION IN SOCIAL FINANCE

Taken in its broad sense, social finance relates to funding into enterprises or initiatives that aim to secure social, environmental and, in the case of impact investment, financial returns.[27].

There are different types of funding mechanisms that come under the umbrella term 'social finance'. Outcomes-based contracting is one such mechanism. When a program is funded through outcomes-based contracting, payments are linked to performance as measured by clearly defined outcome indicators.[28] Compared to traditional funding agreements with social impact service providers, different forms of outcomes-based contracts provide different levels of risk for funders and service providers. Traditional agreements are usually acquitted through a report on expenditure (i.e. activities or services delivered), whereas reporting for outcomes-based contracting is focused on achievement of agreed outcomes. This means that there is lower accountability for *how funding is used*, but greater accountability for *what is achieved.*

In impact investment, an investor seeks both social impact and a financial return. This may occur for example when an investor provides loan or equity finance to a social business. In addition to the outcomes measurement and evaluation of social impact, there is the additional outcomes measurement of financial return and the balance of social impact and financial return.

Outcomes measurement and evaluation in social finance settings follows the same general process as outlined within the *Roadmap*. However, it is important to consider how the nature of contracting arrangements may affect your outcomes measurement and evaluation work. In particular, **you will need to think carefully about**:

- Making sure you balance the priorities of investors with those of the beneficiaries.
- The extent to which you will consider the nature of the funding agreement within key evaluation questions.
- Ensuring you have access to high quality data to demonstrate achievement of outcomes.
- Any additional resources you will need.

POSITIONALITY IN OUTCOMES MEASUREMENT AND EVALUATION

Evaluators bring their own backgrounds, values, and biases to their work, which can influence all stages of outcomes measurement and evaluation. 'Positionality' refers to understanding and reflecting on these influences to ensure that evaluation processes are fair and inclusive.[29] Depending on stakeholder needs and preferences, it may be useful to reflect on positionality and values as part of establishing your purpose for outcomes measurement and evaluation.[30],[31] This can help with open discussions around priorities for the design of your outcomes measurement and evaluation work.

Case study: The Aspire social impact bond

The Aspire program involved providing intensive case management to people who are homeless, including rapid (re)housing and wraparound post-housing supports to increase the chances of people staying housed. The Aspire social impact bond program was financed through private investors through a social impact bond, which is a form of impact investing and outcomes-based contracting. Aspire was evaluated in 2022 by the CSI nodes at Flinders University and the University of Western Australia.[32]

The Aspire evaluation took a mixed methods approach that involved analysis of quantitative datasets (included linked administrative datasets) provided by a range of government agencies and service providers, and in-depth interviews with Aspire participants, staff, key stakeholders and investors. It examined components of the program relating to process, outcome, innovation and investors. In addition, the evaluation undertook an examination of the cost offsets of the program (i.e., the extent to which the program results in reductions in government service costs such as hospital and justice costs).

Reporting against the process of implementation and its effectiveness in improving homelessness outcomes involved, amongst other process and outcome measures, assessing performance against appropriate indicators and calculating investor returns. An additional focus of this evaluation was to explore the role of the social impact bond in the effectiveness of Aspire. This occurred through analysing data with respect to how the social impact bond may have addressed or enabled program flexibility, risk and performance. The Aspire social impact bond evaluation demonstrates just some of the additional considerations for conducting outcomes measurement and evaluation in social finance and philanthropy settings.

2

PLAN

Before diving into outcomes measurement and evaluation, **you need to ensure that your organisation's strategy, culture, stakeholder engagement, human resources and budgets are set-up to support this work (or build them!).** Your organisation should have an established **culture of learning,** and understand the **importance of outcomes measurement and evaluation**.

FOSTERING A CULTURE OF LEARNING

Outcomes measurement and evaluation do not happen in a vacuum; it requires an organisation that is ready, willing, and able to engage in continuous learning**. A 'culture of learning' is important for staff to adopt mindsets that are conducive to using evidence to inform decisions about programs – whether that be for use by internal or external stakeholders**. If a culture of learning is present, this will help to establish the conditions needed for outcomes measurement and evaluation to best serve their key purposes (which are for program development, accountability, improvement, contribution of new knowledge and support for innovations).

An organisation with a learning culture engages in self-evaluation, self-reflection, and self-examination.[33] It considers the impact it is seeking to achieve, takes responsibility for it and actions results to challenge or support its activities.[34] It values candor, challenge and genuine dialogue, with staff being able to use language relevant to outcomes measurement and evaluation. A strong culture of learning supports experimentation and risk-taking and values the opportunity to learn from mistakes. Results from outcomes measurement and evaluation are visible on meeting agendas, in annual reports, on the website, on social media platforms, in performance reviews and throughout communications and reporting materials with partners or funders. The leadership team lead by example, building capacity for, and investing in outcomes measurement and evaluation, while being held accountable for the results.

How to build a culture of learning

Understand your organisation's position: Self-assessment tools (see Appendix 1) can help assess the extent to which outcomes measurement and evaluation are embedded in your organisation, inform action planning, and be used to monitor progress.[35] Used across organizations, such self-assessment tools can be conversation starters in internal engagement processes aiming to create or strengthen cultures of learning.

Leadership: A guiding coalition of champions, participants, influencers, change agents, and communicators lead a strong learning culture. The Board, CEO and Executive should be champions and provide structure including support systems, clear roles and responsibilities, performance reviews and reporting mechanisms. Their own reporting and accountability processes should be results-led. Leadership should also promote outcomes measurement and evaluation when engaging with partners and in public forums. Assess the skill set of your Board - ensure there is someone with outcomes measurement and evaluation expertise who will advocate for their use in ongoing learning and ask relevant questions.

Systems: Assess your current policies, procedures, data management systems, and accountability plans to see if they align with and support outcomes measurement and evaluation – for both internal and external purposes. Does infrastructure (such as IT platforms) need to be developed? Is planning documentation in order? Are there ways to integrate with existing data collection and reporting systems? Are there procedures in place within partnerships and collaborations for outcomes measurement and evaluation to be appropriately conducted and used? What resources will be required?

Capacity, capability, and connection: What capacity and capability exists, where and in who? What are the professional development needs? Assess and offer training or access to new knowledge. Consider connections including networks (such as Social Impact Measurement Network Australia, professional associations such as the Australian Evaluation Society, service networks or peak bodies), partnerships, mentors, and universities (academics, students).

Learning orientation: Outcomes measurement and evaluation is ultimately about learning and improvement. Your organisation should build opportunities for learning through communication loops, regular discussion (such as at team meetings), training, mentoring, and conferences. Insights need to be understood in terms of what is and is not working. This learning needs to be acted on by stopping, growing or embedding particular ways of working. Who will decide which action is to be taken? How and by who will these actions be monitored? The outcomes measurement and evaluation loop should be cyclical and ongoing!

Merit and quality: Understand what merit and quality look like for your outcomes measurement and evaluation system. **Quality** means the outcomes measurement and evaluation system connects with your organisation's mission and values, and is based upon integrity, respect, responsiveness (adaptation based on results), stakeholder involvement, transparent communication and cultural responsiveness.[iii] **Merit** means:

- Applying established and appropriate methods.
- Focusing on all the types of outcomes and impacts created (positive, negative, un/intended).
- Attribution (claiming only the difference you know you've made).
- Utility (how the findings can be applied).[36]

iii The principle of culturally responsiveness refers to the practice of paying particular attention to social and cultural factors in interactions with people from different cultural and social backgrounds.[36] People working in culturally responsive settings continuously self-reflect to examine and unlearn their own assumptions about race and culture.[37]

Case study: Financial Inclusion Action Plan (FIAP) Evaluation

The Financial Inclusion Action Plan (FIAP) evaluation[39] showcases a collective, systems-wide approach to fostering a culture of learning. Run by Good Shepherd, the program enabled 30 participating organisations across financial services, government, and community sectors to take meaningful action towards financial inclusion and resilience. CSI at the University of New South Wales (UNSW) conducted both an impact evaluation and a process evaluation as part of the FIAP program. Through these evaluations, they contributed to building a culture of continuous learning in the following ways:

Impact Evaluation:

- CSI UNSW developed a baseline and indicator bank, providing organisations with tools to measure their progress toward intended outcomes.
- This baseline served as a foundation for participating organisations to track their short, medium, and long-term progress and adjust their strategies based on evidence, promoting iterative learning and improvement.

Process Evaluation:

- CSI UNSW examined what worked (or didn't), for whom, and why, allowing organisations to learn from the early implementation of the program.
- By identifying challenges and opportunities, the process evaluation provided actionable feedback to improve subsequent phases of the program.
- The evaluation also facilitated reflective practices within participating organisations, helping them understand the factors behind their successes or setbacks.

These evaluations directly supported the program's culture of continuous learning by embedding reflection and adaptability into its framework. By sharing findings and facilitating knowledge exchange between participating organisations, CSI UNSW ensured that learning became a core part of the FIAP program's DNA.

Integrating evaluation and learning into every phase of the program ensured that the FIAP Program was not only impactful but also adaptive and future focused. The program demonstrated how building a culture of learning through structured evaluation and collaboration can lead to sustained organisational and systemic change.

KNOW YOUR PEOPLE: STAKEHOLDER ANALYSIS

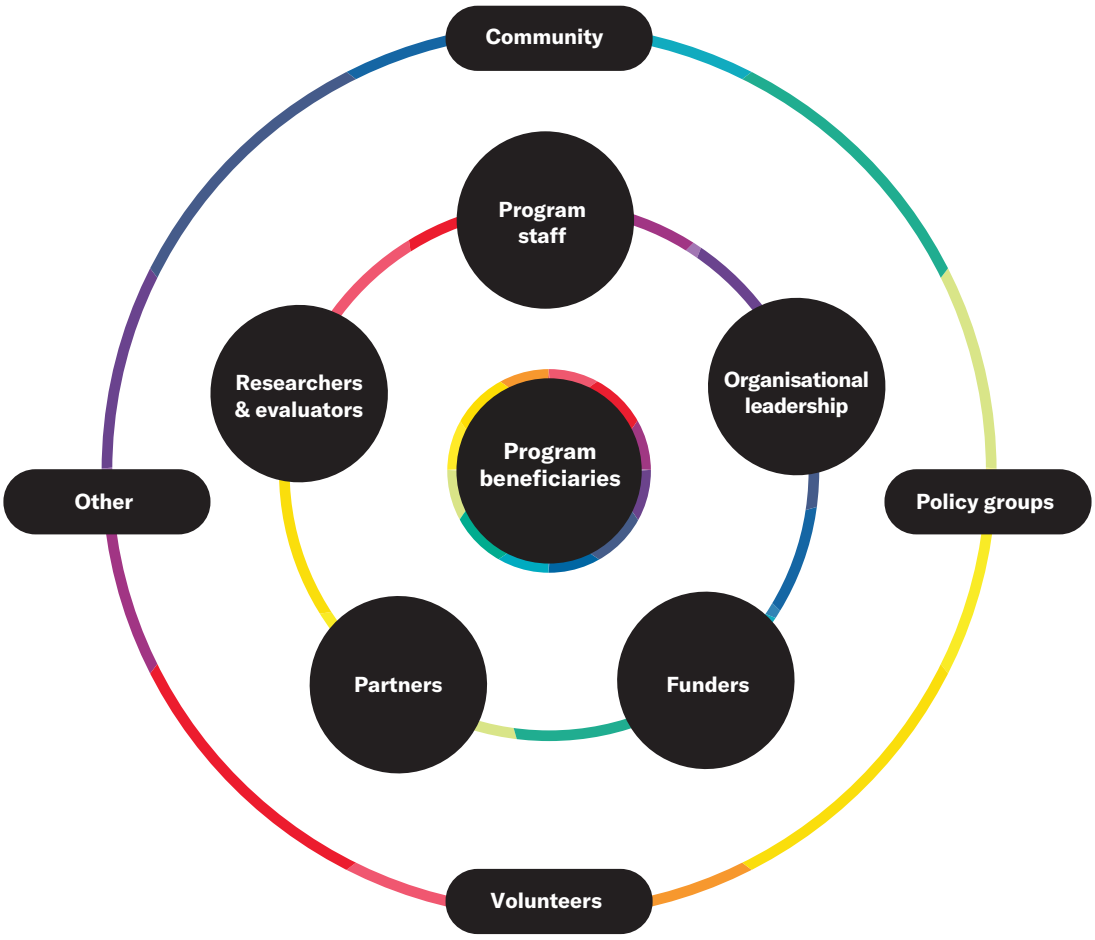
Outcomes measurement and evaluation are more likely to be successful if your stakeholders are engaged with the process. The first step is understanding who your stakeholders are, their level of interest in and influence over the program, and their attitudes towards, and aptitudes for, outcomes measurement and evaluation.

Identifying and analysing your stakeholders

The diagram below (Figure 3) provides an example list of internal and external stakeholders who might be relevant to your organisation or program. While systems thinking (outlined in the previous section) allows you to map all players (micro, meso and macro) who have influence on the outcomes of an issue being addressed, stakeholder analysis focuses on those organisations or groups that intersect directly with your program, and therefore can contribute meaningfully to understanding the difference your program has made.

Figure 3:

Centrality of stakeholder groups relative to outcomes measurement and evaluation



Alongside their roles in the program, the priorities, interests, and needs of your stakeholders for outcomes measurement and evaluation need to be understood. Consider how they will be impacted by your program, what is needed to support their meaningful involvement, what success looks like from their perspective and any risks that need to be considered.[40] Some stakeholders may seem peripheral, yet important to engage. Identify which stakeholders might have ***resistance*** to what you are trying to achieve (and why they may be resistant, and how to address their concerns), how to increase ***engagement and/or ownership*** (and how to sustain it). Outcomes measurement and evaluation may be met with resistance due to:

- Poor stakeholder engagement in initial program design or as part of outcomes measurement or evaluation.
- A lack of capacity (time and resources) to measure outcomes and evaluate, especially within smaller organisations.
- Lack of expertise in how to measure outcomes and evaluate.
- Lack of understanding about the importance of outcomes measurement and evaluation.
- Lack of leadership support.
- An existing perception that a program is working despite limited or lack of evidence.
- Thinking that clients may disengage if they are asked to participate in outcomes measurement and evaluation.[41]

For example, in the *Sport* case study, there might be resistance to outcomes measurement and evaluation from staff implementing the program as they have low skills in data collection and find it a burden, or if they have not been involved in program design from the outset and think adjustments are needed. Yet, they may become champions if they are engaged in outcomes measurement and evaluation early on (including when clarifying the program design), trained and provided with the tools to measure, are provided with early access to results, and asked to help interpret the indicative results, if they understand the benefits of measuring the impact of their work and how it may help them improve outcomes for young people.

From your stakeholder analysis, it is important to think about the level of engagement appropriate for each stakeholder. This can be:[42]

- **Passive:** no engagement, no communication, no relationship.
- **Monitoring:** one-way communication, no relationship.
- **Informing:** one-way communication, short or long-term relationships.
- **Transacting:** work together in a contractual relationship.
- **Consulting:** information is gathered from stakeholder for decision making.
- **Co-design:** work directly with stakeholders to ensure their concerns are considered in decision making.
- **Collaborating:** mutually agreed solutions and a joint plan of action is delivered in partnership with stakeholders.
- **Empowering:** decision-making is delegated to stakeholders.
- **Ownership/self-determination:** stakeholders set priorities, make decisions and pursue their own development on their own terms[43].

For example, Table 2 (on next page) shows an example of stakeholder mapping, role description and needs by group.

Table 2:

Example of stakeholder mapping, role description and needs

Stakeholder groups	Examples of role description and contribution to outcomes measurement	Examples of needs	Example of priority level
Measurers (e.g. practitioners; researchers; evaluators; consultants)	Design and implement outcomes measurement and evaluation frameworks, analyse and present the results	Resources; an understanding of needs and scope; buy-in from key stakeholders	Medium
Program managers and funders (e.g. government; corporates; NGOs; enterprises; business)	Possible funders of outcomes measurement and evaluation; inform design; provide data if relevant; use evidence from outcomes measurement and evaluation to write policy; to inform future investment; to develop or refine initiatives	Relevance; accessible, useful, informative information; engagement	Medium - high
Beneficiaries (clients/ consumers) and communities	May be partners in the design and implementation of outcomes measurement and evaluation; primary providers of knowledge and data	Respect and trust; participation & engagement; accessible information; feedback	High
Innovators & implementers (e.g. social purpose organisations; communities)	Inform design (may be conducting or partners in the design); data collectors; data providers and reporters	To know if they are meeting their purpose/ objectives; buy-in to the design & data collection; accessible, useful, relevant information	High
Others (e.g. partner organisations; service providers; advocacy groups; family members; etc)	Data providers; may inform the design	Engagement; accessible information; relevant feedback	Low – medium (depending on the group)

Sourced from the Compass.[44]

Think about the parts of the outcomes measurement and evaluation process with which your stakeholders will be involved in – relating to planning, question development, defining indicators of success, design, data collection, review, knowledge translation and action plans. Think also about the control they have over these processes. Social impact work takes place in a political context. The political context is especially important to understand in outcomes measurement and evaluation as this often determines the reallocation of resources, how the needs of vulnerable groups are prioritised, how different stakeholders with complex relationships are engaged.[45]

Your stakeholder analysis should support your understanding of relationships and politics surrounding your program.

Issues of budget, human resourcing, access, equity, decision-making processes, sustainability, capability relating to outcomes measurement and evaluation, and organisational capacity for stakeholder engagement all need to be considered as part of your stakeholder engagement strategy.

THE ROLE OF LIVED EXPERIENCE IN OUTCOMES MEASUREMENT AND EVALUATION

As **program beneficiaries usually represent a high priority stakeholder group, it is good practice to consider their lived experience as central to your outcomes measurement and evaluation process** as shown in Figure 3. This reminds us that outcomes measurement and evaluation are about ensuring best practice and improved outcomes for the community's benefit. It often links to mission and values of supporting voice and citizenship, respecting human dignity and worth.[46]

Lived experience refers to the firsthand knowledge that individuals gain through direct involvement with social issues or programs. Incorporating lived experience into outcomes measurement and evaluation processes is essential for ensuring that evaluations are grounded in the realities of those most affected by programs.[47],[48],[49],[50] Some key benefits of engaging people with lived experience include:[51],[52],[53]

- **Relevance**: People with lived experience provide unique insights into how programs affect them directly, offering context that may not be captured through traditional data collection methods.
- **Ethical engagement**: Engaging people with lived experience ensures that outcomes measurement and evaluation respects the dignity and agency of beneficiaries, fostering more ethical and culturally responsive practices.
- **Enhanced credibility**: Outcomes measurement and evaluation that incorporates lived experience is often seen as more credible because it reflects the voices of those most affected by programs.
- **Broader impact**: By involving people with lived experience, you can increase stakeholder buy-in and ensure that findings lead to actionable change.
- **Improved outcomes**: Programs that incorporate feedback from those with lived experience are more likely to achieve meaningful outcomes because they are informed by real-world challenges and opportunities.

Mechanisms for engaging beneficiaries with lived experience in outcomes measurement and evaluation include:

1. **Governance and oversight**: Engage individuals with lived experience through appropriate governance and oversight. This ensures that appropriate decision making is in place for outcomes measurement and evaluation design, implementation and knowledge translation.
2. **Advisory or reference groups**: Establish advisory or reference groups composed of people with lived experience who can provide ongoing feedback throughout the outcomes measurement and evaluation process.
3. **Co-design:** Engage individuals with lived experience in co-designing outcomes measurement and evaluation frameworks and tools to ensure their design is relevant and respectful.
4. **Participatory data collection and analysis**: Use participatory evaluation methods (e.g., co-facilitated focus groups or community consultations, collaborative sense-checking of data) that allow individuals with lived experience to actively contribute to data collection and analysis.[54],[55],[56]
5. **Co-authorship:** Co-authoring ensures that appropriate acknowledgement is given to those who have provided knowledge and insights for outcomes measurement and evaluation findings.

Consult with appropriate bodies when working through how to engage with lived experience representatives. For instance, there may be different considerations, needs, risks and protocols that apply to working with young people and children, people culturally diverse communities, people escaping domestic violence or refugee populations, for example. Advocacy groups, peak bodies or not-for-profit specialist organisations who work with specific groups can usually provide excellent guidance (and contacts for partnering) on how to engage appropriately.

It is also important to ensure that evaluation processes are culturally safe by respecting diverse perspectives and addressing power imbalances between evaluators and participants. For example, Indigenous-led evaluations prioritise self-determination, where Indigenous communities co-design and lead evaluation efforts (see below for an overview of cultural safety).[57,58]

Organisations need to reflect on and be willing to change their structures, mindsets and communications, as well as provide support and training to beneficiaries to facilitate meaningful participation. Understanding who stakeholders are, how they interact with each other and the program, their histories, cultures, attitudes towards and need to be involved in outcomes measurement and evaluation will not only support the delivery of the program but also lead to more relevant and rigorous outcomes measurement and evaluation.

Case study: Integrating lived experience for research into crisis accommodation in Australia

A research report titled, 'Crisis accommodation in Australia: now and into the future'[59] and examined the key elements of effective and appropriate crisis accommodation models. To do so, the research team engaged with individuals with lived experience of crisis accommodation, as well as frontline staff and stakeholders across Australia.

A total of 21 interviews were conducted with people who have experienced crisis accommodation firsthand. The interviews focused on participants' experiences during their stays and provided valuable feedback on how these services can be improved. Although the sample size was small, the perspectives gathered offered critical insights into the functioning of crisis accommodation in Australia. Participants were recruited through a specialist homelessness service in Melbourne and two newly established homelessness alliances in Adelaide. Case managers helped approach potential participants, ensuring voluntary participation without impacting their services. Interviews were conducted in public locations such as service offices and cafes, and participants received a $50 supermarket gift voucher as a token of appreciation.

Including feedback from individuals with lived experience is essential to understanding the strengths and weaknesses of the system, and to identifying how it can be improved. Participants in the study emphasised the need for a comprehensive range of support services in crisis accommodation, extending beyond just housing. Many felt that services should address physical health needs, provide material aid (e.g., food, bedding, clothing, transport tickets, and phones), offer support for substance use, assist with Centrelink issues, provide legal assistance, and offer help with child protection. The findings also highlighted the importance of expanding the lived experience workforce, particularly by increasing the recruitment of Indigenous workers and raising awareness about the value of these roles. In response, the service has integrated lived experience practices into its operations and is working to involve young people with lived experience in governance.

OUTCOMES MEASUREMENT AND EVALUATION WITH INDIGENOUS PEOPLE AND COMMUNITIES

Cultural safety

In Australian Indigenous settings, cultural safety refers to Indigenous people feeling like their cultures are valued, their experiences are validated and their knowledges and skills are recognised. Cultural safety also refers to absence of all forms of racism in Indigenous peoples experiences of programs, policy development, evaluation, research and service delivery.[60]

During the planning phase of outcomes measurement and evaluation for programs affecting Indigenous people, **consider how you can embed Indigenous perspectives of cultural safety within your work**. This might be through ensuring that Indigenous programs participants can provide feedback on their experiences of cultural safety. It also might be through having Indigenous stakeholders and staff in relevant roles such on governance committees, as project leads and as evaluators.

INDIGENOUS DATA SOVEREIGNTY

Indigenous data sovereignty[61] refers to control of local Indigenous people over the design, collection and use of data in outcomes measurement and evaluation. This principle is put into practice through **decision-making by appropriate local Indigenous authorities using data disaggregated at the local level**. This puts self-determination into practice and ensures benefit and use of outcomes measurement and evaluation are maximised. Indigenous data sovereignty also ensures that local communities have the opportunity to make sure outcomes measurement and evaluation processes and findings are contextualised and relevant.

Indigenous data sovereignty occurs in practice through similar mechanisms to those outlined above for engaging people with lived experience in outcomes measurement and evaluation, but the emphasis is on *local decision-making and control* over outcomes measurement and evaluation. For example, it would not be enough to get suggestions from a Indigenous Reference Group about how to engage Indigenous participants in an evaluation. Indigenous representatives who have local expertise and relevant authority must also be involved in clarifying mechanisms of change for a program (including identifying outcomes) and designing key evaluation questions as well as designing methods.

Case study: First Nations Cultural Safety Framework

The Australian Evaluation Society's First Nations Cultural Safety Framework[62] aims to provide guidance for evaluation practitioners, policy makers and other stakeholders on how to implement cultural safety within of evaluation practice. It:

1. Outlines principles of culturally safe evaluation.
2. Provides practical guidance on the roles and responsibilities of different stakeholders in the evaluation process.
3. Provides practical guidance on what contributes to culturally safe evaluation in all phases of the evaluation process.
4. Identifies the outcomes that can be achieved through implementation of the Cultural Safety Framework.

The framework describes critical self-reflection, cultural accountability, addressing and preventing racism and understanding systems of power and privilege as key practices of cultural safety. It describes several examples and poses reflection questions for evaluators to consider their roles and responsibilities in practicing cultural safety in evaluation. Working towards culturally safe evaluation is explained as an ongoing commitment and journey.

The First Nations Cultural Safety Framework is a key document that should be used to guide evaluations that affect Indigenous people and communities.

ETHICS

Research ethics involves the application of fundamental ethical principles to research and evaluation (including outcomes measurement). Applying these principles helps to ensure:

- **Merit**: i.e., research or evaluation questions are worth asking.
- **Integrity of methodology**: i.e., research or evaluation questions can be answered by the intended approach.
- **Psychological and physical safety** for participants and researchers or evaluators.
- **Respect for time and voice**: i.e., findings are utilised and interpreted with fidelity.

During the planning phase of your outcomes measurement and evaluation work, it is important to familiarise yourself with ethical principles for human research and outcomes measurement, through the following frameworks:

- The NHMRC's National Statement on Ethical Conduct in Human Research (2023).[63]
- AIATSIS's Code of Ethics for Aboriginal and Torres Strait Islander Research (2020).[64]

You may need to apply for ethics approval from a recognised Human Research Ethics Committee before you begin your outcomes measurement and evaluation work. Different contexts and institutions have different formal requirements for ethics approval processes. You may need to submit all of your planning documents as part of your application (including outcomes measurement and evaluation framework and data collection tools, etc), which is why it is important to look into ethics requirements and deadlines when you are planning. For example, it may take several months for the ethics approval process to be completed so you will need to incorporate this into your timelines and budget. You should follow the appropriate process depending on relevant requirements and be guided by key documents outlined above.

Applying ethics principles to outcomes measurement and evaluation may require slightly different approaches. See the Australian Evaluation Society's (AES) Guidelines for the Ethical Conduct of Evaluations[65] for specific guidance on ethics in evaluation.

Further guidance is provided about ethics in Section 6.

BUDGETING FOR OUTCOMES MEASUREMENT AND EVALUATION

At the beginning of a program, it can be hard to know the resources you will need to measure your outcomes and evaluate. For this reason, you might need to come back to this step after you have developed a more detailed evaluation plan. Or, you may have an allocated budget that you need to use to guide your outcomes measurement and evaluation work.

To understand the budget you will need, you first need to understand what data you will need to collect, the frequency of data collection, number and type of stakeholders you will collect data from and the type of evaluation you want to complete. You also need to decide whether outcomes measurement and evaluation will be an in-house or external activity. If you do not have much control over the budget allocated to outcomes measurement and evaluation, you will need to decide the suite of approaches that you can afford to help measure your outcomes and evaluate. Consider:

- The purpose of your outcomes measurement and evaluation work.
- The significance of the findings, for example contributions to program decisions.
- The requirement for quality evidence by stakeholders.
- The likelihood of scrutiny.
- The needs of your key stakeholders.
- The amount of work required to be clear on the scope of your outcomes measurement and evaluation work.
- The type of evaluation you are planning to do.
- Whether you need one, or more, data sources (e.g. survey and in-depth interviews).
- Alternative methods of data collection: face-to-face (more expensive), telephone, mail, online (this will also depend on the characteristics of your potential respondents), other systems-based or participatory evaluation activities.
- Alternative sources: administrative and secondary data, other organisational data readily available.
- Timing of data collection (Could some additional information be collected at in-take?).
- Roles and responsibilities for outcomes measurement and evaluation tasks (Should data collection and analysis be done internally or externally e.g. would training of internal staff be cost saving in the long-term?).

Outcomes measurement and evaluation can be simple or sophisticated, depending on the complexity of the setting in which you are working, access to resources and requirements for quality evidence. For example, Randomised Controlled Trials (RCTs) are considered a 'gold standard' design for impact evaluations as they provide the basis for outcomes measurement of what would have happened in the absence of the program so providing a reference point against which the difference the program makes can be measured. However, they are not cheap to conduct, are difficult to implement in the social program context and do not always achieve stakeholder adherence often due to perceived ethical concerns. Additionally, these types of designs must be implemented through a process that involves key stakeholders appropriately to ensure their relevance and appropriateness.

Very often, available resourcing will determine the scope of the evaluation. **As a guide, around ten percent of the program budget should be allocated towards evaluation.** If high quality evidence is needed, this may increase to around twenty percent. On the other hand, if a program is smaller and the findings are likely to be less significant, then you may be able to allocate closer to five percent of your budget towards evaluation. It is important at this early stage to determine how much resourcing is available for outcomes measurement and evaluation so that you can plan your approach and design accordingly.

You will need to plan for resources. If appropriate, allow time for staff to be trained in outcomes measurement and evaluation and time for the actual planning, data collection, data analysis and knowledge translation activities (which will vary with the type of outcomes measurement and evaluation methods). You might need to employ additional staff to support your evaluation needs. Money is an important resource. Cost planning is speculative, and it is essential to allow for contingencies[66]. You should base your cost estimate on previous experiences, expert advice, and thorough planning.

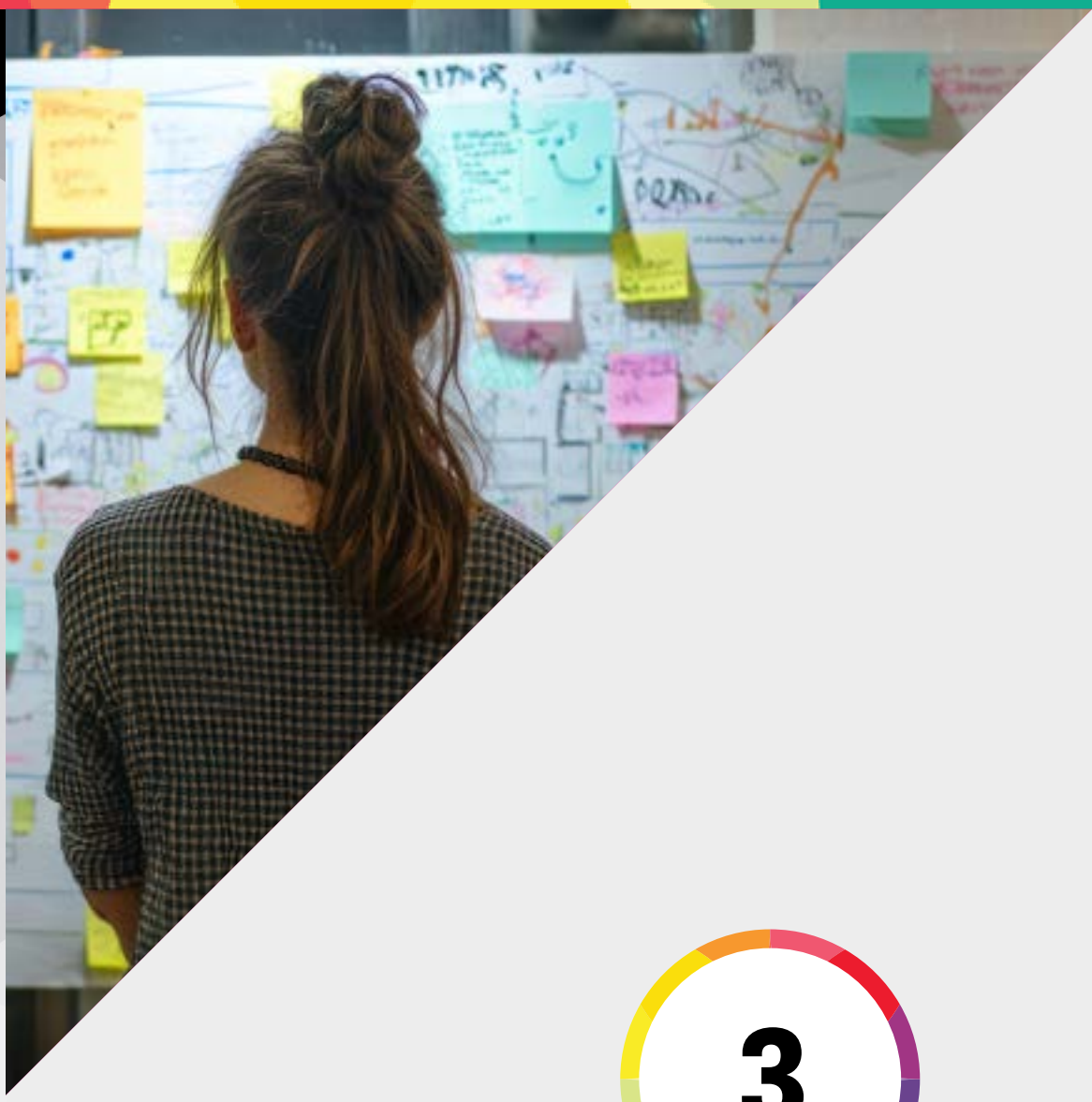
The risk of under budgeting for outcomes measurement and evaluation is high. Risks associated with this include inability to capture all outcomes and misrepresentation of program achievements. Not allocating sufficient resources (staff, time and money) towards knowledge translation activities can make outcomes measurement and evaluation pointless, through missing out on the opportunity to engage relevant stakeholders and implement change[67]. There is a range of free resources to support organisations looking to complete outcomes measurement and evaluation. Some of these are listed in Appendix 1.

INTERNAL AND EXTERNAL EVALUATIONS

Evaluation and outcomes measurement can be conducted externally by engaging a qualified researcher or evaluator, or internally, by your skilled staff. There are advantages and disadvantages to using internal evaluators, including the benefits of relationships with staff and stakeholders whereby they can support and help to navigate engagement, and represent different perspectives[68]. Additionally, internal evaluators usually have more in-depth knowledge of the program context. In terms of disadvantages, there may be questions about the credibility of the evaluation and bias. External evaluators may be seen as producing more credible results due to their independence from a program. External evaluators may also have specialised skillsets in particular approaches and tools.

In summary, in planning for your outcomes measurement and evaluation work it is important to consider:

- Who your stakeholders are, how they can participate and how you will engage them.
- The specific needs of your beneficiaries or communities who have a stake in the evaluation.
- Ethical considerations in terms of the scope, questions asked, engagement approaches and use of findings.
- What an appropriate budget would be (or what budget you have available).
- Whether to commission the evaluation externally or use internal resources and expertise.



3

CLARIFY PROGRAM DESIGN

TOOLS FOR MODELLING CHANGE

An important foundational activity for outcomes measurement and evaluation is to model what changes you are hoping to achieve through your program, and how and why activities lead to these outcomes. Two common tools for undertaking this are theory of change and program logic[iv]. If done well, they are critical to bringing clarity to outcomes measurement and evaluation work. Other benefits include demonstrating rigour (both theoretical and practical) in process as well as outcomes and enabling clear communication internally as well as to external stakeholders.

Theory of change

A theory of change is **an explicit theory or model of how a program will achieve the intended or observed outcomes**.[69] It articulates the hypothesised causal relationships between a program's activities and its intended outcomes and identifies how and why changes are expected to occur. In doing so, the theory of change comprises a change model (the changes the program intends to achieve) and an action model (the activities that will lead to those changes).[70] **A theory of change must be plausible, doable and testable.**[71] It should also articulate the assumptions and enablers that explain why activities will lead to the outcomes outlined. While a theory of change is often represented as a diagram or chart, a narrative can also be used.

A theory of change will help you understand how your program will achieve its goals. It will help you with:[72],[73]

- **Strategy:** Helps teams work together to achieve a shared understanding of a program and its aims; ensures all activities align with the purpose of the program; encourages in-depth thinking about the program and its assumptions.
- **Outcomes measurement and evaluation:** Helps to formulate and prioritise evaluation questions, identify outcomes and plan the outcomes measurement and evaluation process; encourages the use of existing evidence.
- **Communication:** Informs stakeholders, in an 'elevator pitch'- type of approach, about the program's aims.
- **Working in partnership**: When programs are delivered in collaboration, developing a theory of change will help clarify roles and responsibilities.

To formulate your theory of change, start by defining the main activity for your program and its long-term outcomes. These represent the 'start' and 'end' of your theory of change (what you do and for what purpose). Clearly outline the change model (the changes that will result from your program). You can then articulate the main processes or activities (the action model) through which you engage with your target group, population, or community to achieve those outcomes. Your theory of change should be informed by evidence of 'what works' to address the problem you are seeking to solve (e.g. similar programs or approaches in different circumstances), or evidence that an innovative approach (e.g. engaging with groups at different times, in different circumstances) is likely to work and why.[v]

You should also consider the enablers that support you to deliver your program and achieve your goals. Internal enablers are conditions or factors that need to be in place for your program to work and are mostly within your control (e.g. governance, relationships, quality of services). External enablers are factors outside your immediate control and describe the environment in which your program operates (e.g. social, cultural, political, economic factors). See Figure 4 (on next page) for the theory of change for *Sport*.

---

iv Theory of change and program logics are just two tools that you can use to show change mechanisms. Other non-linear tools such as program models can also be used.

v Some theories of change also discuss assumptions (why activities are assumed lead to certain outcomes, or why an intermediate outcome is assumed to lead to a long-term outcome), but we include these in the program logic, which further details the theory of change.

Figure 4:

Theory of change - fictional case study, *Sport* program

Sport

Enables primary school children to engage in sport activities and learn about healthy living habits

By providing information sessions and resources for students, teachers and parents to promote healthy living habits and opportunities to engage in physical activities 2-3 times per week

Information sessions for teachers to promote healthy living habits and how to include these in the classroom routine	Information sessions for parents to promote healthy living habits relating to sleeping, eating and exercise (biannual)	Information sessions for students to promote healthy living habits (once per term)	Take-home resources like flyers and fridge magnets as reminders of healthy living habits	After school free of charge sport programs for every student (allowing for each child to participate 2-3 times per week)

To improve attitudes towards and likelihood to lead healthy active lives

Students learn how to lead healthy lives	Students increase their levels of physical activity	Students improve their eating and sleeping behaviours	Students are healthy physically and mentally

Such that healthy living habits become the norm, leading in the long-term to positive health outcomes into adolescence and adulthood

Enablers: Parents allow children to participate in the school-based sport activities; Teachers are supportive of the program and the school allows use of school grounds

Tips on developing a theory of change

1. Organise a workshop. Start with your team and consider including other key stakeholders.
2. Prepare flip-chart paper, post-it notes and textas. Or use an interactive, visual online tool, like Miro Board. Explain the components of a theory of change (activities, long-term outcomes/ goals, enablers).
3. Use different colour post-it notes for each category. Start with one category - usually starting with the goals as most people will have a good idea of what they want to achieve; ask everyone to write the program goals or long-term outcomes on a post-it note. Place those at the bottom of the flip-chart paper.
4. Ask everyone to discuss what outcomes (intermediate or longer-term) need to be achieved to reach this goal and what activities will support the achievement of those outcomes. Place the activities at the top of the paper and any intermediate outcomes in the middle.
5. Create visual representations that show causal pathways for how each activity leads to change. Take time to discuss, remove duplicate ideas/ concepts, rearrange for timelines and, relevance, and add enablers. Outcomes will be based on assumptions (what participants think will be achieved based on experience, or current evidence). Make sure you take note of these to include them in the program logic when you expand on the theory of change.
6. When you are confident with the draft theory of change, circulate it to other stakeholders and ask for feedback.

Remember the 'Ikea effect'[74] - people relate more to and have a greater commitment and ownership over things they helped to create!

Program logic

A program logic is **a visual representation of how your program will achieve its goals, including the short-, medium- and long-term outcomes**[vi] (discussed below). Like your theory of change, your program logic is best developed at the design or planning stage of a program, but if this has not happened, these can be developed, modified and enhanced as the program evolves. Where possible, use evidence to link activities to outputs and outcomes. Outcomes can be based on what you might expect to see (e.g. if students are offered the opportunity to participate in organised sport activities after school we expect many will participate 2-3 times per week and their physical health will improve).

Program logic terminology

Inputs are the necessary resources for a program to run. E.g. staff, volunteers, funding, buildings, technology, machinery.

Activities are what the program is doing and how. E.g. online information, webinars, training, services.

Outputs are numbers or counts of things that result from the program. E.g. number of online webinars, number of participants.

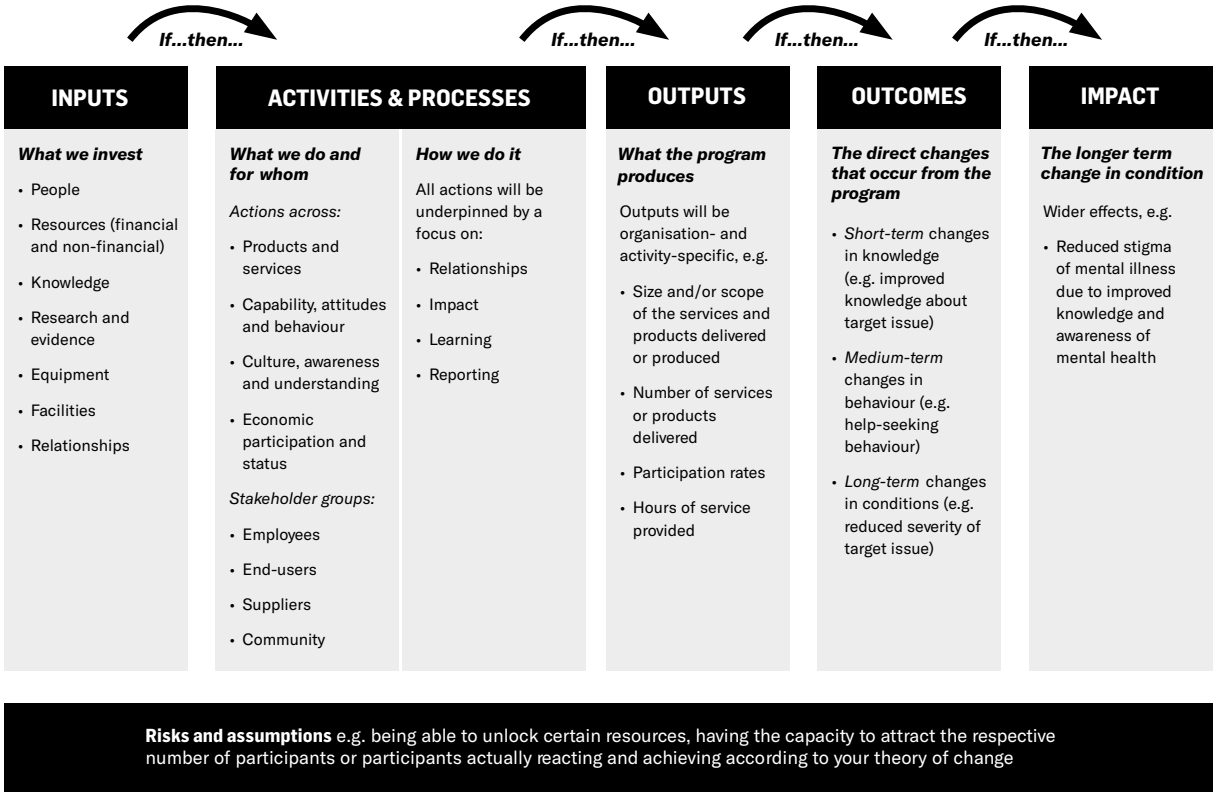
Outcomes are the changes that your program produces in the short-, medium-, and long-term. Long term outcomes may also be termed 'impact'.

Impact is the lasting, systemic change to which your program or organisation contributes.

vi Long term outcomes may also be considered 'impacts'. Note there are also different definitions of 'outcomes' and 'impact', For example, in the health promotion setting 'impacts' refer to the immediate results of a program, whereas 'outcomes' are the longer term changes to health conditions.[76] In the program logic context it is useful to differentiate between 'outcomes' as the immediate and medium-term results of a program such as changes in skills, knowledge and behaviour and 'impacts' as longer term changes in conditions.

Figure 5:

Program logic template

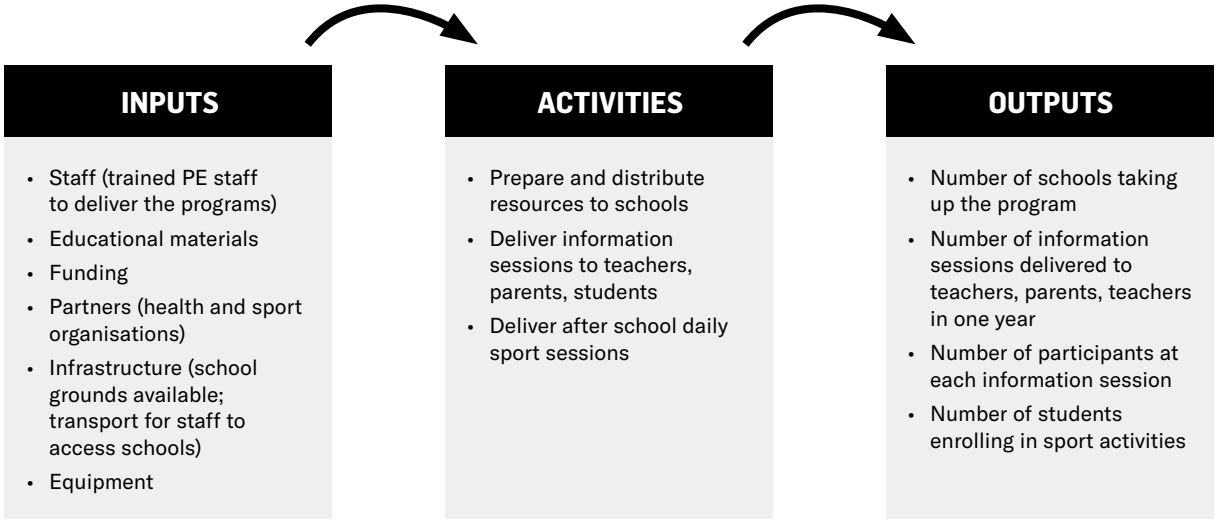


The program logic (Figure 5 above) has an underlying "if-then" relationship, linking a program's necessary **inputs**, **activities**, **outputs**, **outcomes** and **impact** through logic.

Assumptions and risks will accompany your program logic: these are external conditions which could affect the program's progress, but which are not under the direct control of people planning, implementing or managing the program. An assumption is a positive statement of a condition that must be met for the program's objectives to be achieved. A risk is a negative statement of a condition that might prevent the program's objectives from being achieved.[76] You should use evidence (information about other programs, data and experience) to foresee these risks and prepare mitigation strategies.

In the *Sport* example we assume that making after school sport activities freely available to students will result in a higher participation in physical activity. Some factors may interfere with this assumption, for example parents' ability to delay the school pickup, or their adversity towards a respective sport may interfere with students' uptake of the program and realisation of outcomes (risk). Or the risks might be at the school-level, for example lack of infrastructure to support the proposed sport activities.

Figure 6:
Program logic – fictional case study, *Sport* program (Part 1 – inputs, activities and outputs)



Developing the first half of your program logic – identifying **inputs**, **activities** and **outcomes** – relies on your understanding or planning of the program. You should include here the resources necessary for your program to run, from internal support to funding, infrastructure and external partnerships (inputs), the range of activities your program will deliver (activities), and how you will keep track of your delivery of these activities (how much of these activities the program will deliver, how many clients it will engage with). Figure 6 (above) presents the inputs, activities and outputs for *Sport*.

The second part of the logic model, mapping the outcomes of the program, can be more challenging due to the difficulty of identifying outcomes that will result, or confusion that is often experienced about the difference between outputs and outcomes, and outcomes and impact.

IDENTIFYING OUTCOMES

Outcomes - what a program achieves - can be measured at different points in time and at different levels.[77] **Short-term outcomes** capture changes you would expect to see immediately as a result of the program, such as changes in knowledge and skills (e.g. improved knowledge of students about benefits of regular exercise, improved skills of teachers to teach sports).

Medium-term outcomes capture changes that you might hope to see as a result of participating, such as changes in behaviour (e.g. increased participation in regular exercise) and **long-term outcomes** (sometimes called 'impact' – see below) capture changes in conditions (e.g. reduced rates of obesity among school-age children, adolescents and adults).[vii]

When identifying outcomes, you should consider whether your program is expected to yield financial, social, cultural and/or environmental outcomes to ensure you map and measure all potential outcomes. The triple account of outcomes (**social**, **environmental**, **economic**) and targets are often used in accounting techniques (e.g. Triple Bottom Line or Corporate Social Responsibility reporting). When working with culturally diverse groups, it is important to consider which other outcomes may be relevant to that particular group.

vii Sometimes these are referred to as immediate, intermediate and long-term outcomes rather than short- medium- and long-term.[78]

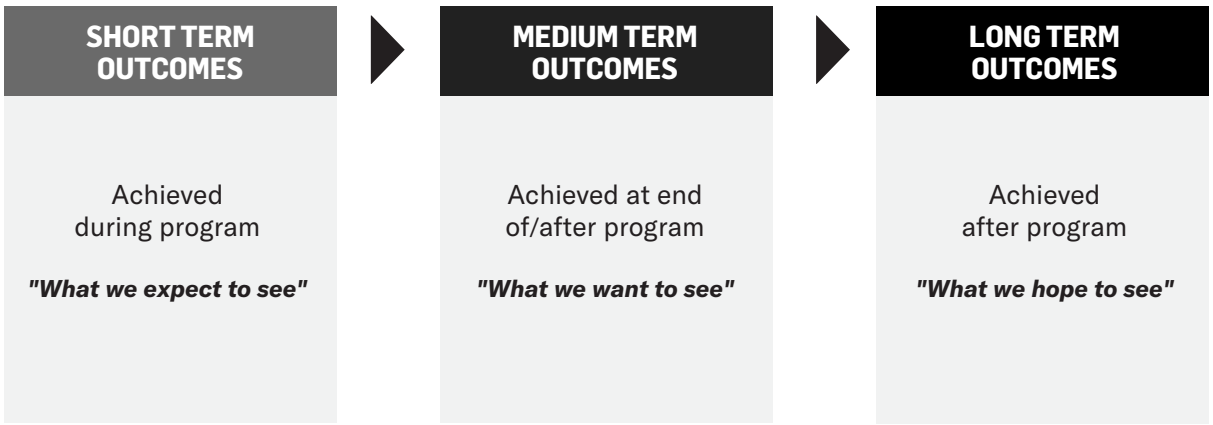
Outcomes can be achieved at individual or program (**micro**) level (e.g. improved quality of sleep); community or organisation (**meso**) level (e.g. reduced crime rate) or at population, industry, or sector (**macro**) level (e.g. reduced hospitalization rates among young adults). While there is no direct link between the timing of an outcome and the level at which it occurs, changes that occur at macro and meso levels are often more complex and require more time to achieve.

Long term outcomes are sometimes called 'impact'. Impact is the long term, systemic-level change your program intends to achieve. Impact relates to the vision of your organisation.

Refer to Figure 7 below for a diagram showing differences in outcome timeframes for short, medium and long term outcomes (adapted from Innovation Network, n.d.[79]).

Figure 7:

Diagram showing outcome timeframes



DEVELOPING OUTCOME STATEMENTS

Once you have identified the intended outcomes of your program you can then develop more explicit outcome statements. **Outcome statements are clear and measurable statements about what your program is hoping to achieve.** They should be specific in stating what will change and by when. Refer to Figure 8 for different types of outcomes and Figure 9 for examples of short-, medium- and long-term outcomes of *Sport*.

There are no definitive guidelines on timelines to measure different outcomes. For example, while medium-term outcomes can sometimes be measurable within a few weeks, in other programs these might only be measured several months or years into the program. Precisely when outcomes can be measured depends on the problem that the program is addressing, the purpose, scope or the target population.

It may be difficult for some programs to measure their long-term outcomes, due to the timeline and complexity of the primary outcome or longer-term impact. For example, it may be years before the long-term outcomes and impact of *Sport* (improved health outcomes in adolescence and adulthood) can be measured, and it is likely that achievement of these outcomes are beyond the control of the *Sport* program anyway. However, you could measure the change in the levels of physical activity (medium-term or intermediate outcomes), which serves as a proxy and may be an indication of the ultimate outcomes.[80]

The extent to which intermediate or medium-term outcomes can serve as proxies is not straight-forward and it requires a thorough literature investigation, consulting organisational data or experienced practitioners. Additional activities or systems-based work may be necessary to facilitate the longer-term outcomes. In some circumstances it may be helpful to also set targets for outcomes - the extent to which change is expected. Targets should be based on evidence and be realistic.

Remember, **not all outcomes are predictable**. It is often hard to project unintended outcomes (positive, negative, or neutral), but these may become obvious as the program matures, and it is important to allow for these to be measured. Collecting qualitative data from a range of stakeholders is a good approach to identify what else is being achieved, in addition to what your model predicted. In mapping potential unintended consequences, you should also think about who else might be affected by your program and the external factors that may influence on your program (e.g. people, circumstances, the environment).

Outcome statements

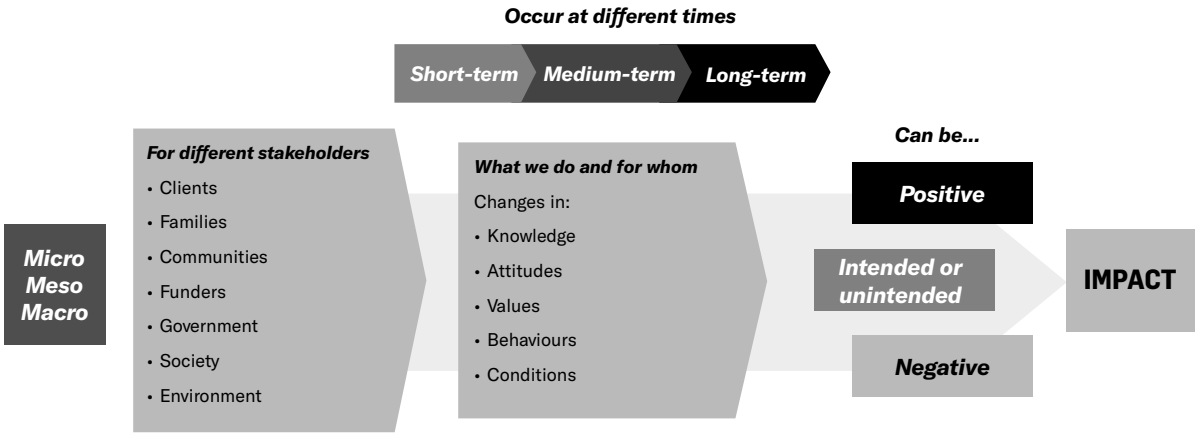
An outcome statement is a statement about who or what will change as a result of a program. Outcome statements:

- ✔ Are specific about who and what.
- ✔ Include language of change (e.g., increasing, more, better).
- ✔ Are attainable/realistic.
- ✔ Are within the scope of influence.
- ✔ Are attributable.
- ✔ Are agreed upon by program stakeholders.
- ✔ Refer to outcomes not outputs (e.g., results not activities).
- ✔ Are measurable.

For example: School students have 25% improved knowledge about healthy living habits by the end of the program in [Year].

Figure 8:

Outcome types



Key points

- Outcomes are not always positive.
- Short- or medium-term outcomes are more directly influenced by the activities of a program. They may be proxies for long-term outcomes, but they will not always lead on their own to long- term outcomes; further activities may be necessary.
- Long-term outcomes are more likely to may also be influenced by factors external to your program.

Types of outcomes

- Short-, medium-, long-term.
- Micro, meso, macro level (check your stakeholder map to ensure all are included).
- Economic, social, cultural and environmental.
- Intended and unintended.

Figure 9:

Program logic – fictional case study, *Sport* program (Part 2 – examples of short-, medium- and long-term outcomes)

SHORT-TERM	MEDIUM-TERM	LONG-TERM
• Students have better knowledge about healthy living habits • Parents have better knowledge about healthy living habits • Teachers have better knowledge about healthy living habits • Teachers are prepared to support children (e.g. by including resources in the daily routine) • Students engage in after school sport activities	• Students engage regularly in after school sport activities • Parents encourage at home healthy living habits: healthy sleeping habits; healthy eating habits; outdoors time	• Students achieve and maintain a healthy lifestyle • Students are less likely to experience life-style related illnesses • Students are less likely to experience mental illness (e.g. anxiety or depression)

Tips to develop a program logic

Developing your program logic is a good opportunity to engage diverse internal and external stakeholders including the evaluation team, people implementing the program, user or client representatives, leadership, funders, etc. You can do this by following a process like this:

1. Use flip-chart paper and post it notes (or an online document everyone can edit). Split the paper (or the online document) into six columns: inputs, activities, outputs, short-, medium- and long-term outcomes.
2. Everyone should write one item on each post-it note (e.g. one input, one output, etc), then place their post-it notes in the relevant column. You may notice some items that you might have thought of as short-term outcomes may actually fit under 'outputs', or that some stakeholders start discussing whether an outcome is medium- or long-term.
3. Shuffle the post-it notes and discuss any points of disagreement or confusion until you have agreed on a program logic that suits your theory of change and program.

When you map the short-, medium- and long-term outcomes it helps to look back at your problem tree and theory of change for a comprehensive picture of the changes that your program seeks to achieve, and your stakeholder map, to ensure you have considered outcomes for all stakeholders (whether engaged in this exercise or not). It helps to begin filling in the short-term outcomes (changes in knowledge), before mapping medium-term outcomes (changes in behaviour) that result from this. The changes in behaviours should point towards changes in conditions (long-term outcomes). Remember that some short-term outcomes may be proxies for long-term outcomes. Consider the dimensions at which outcomes occur (micro, meso, macro) with various groups of stakeholders, and social, financial and environmental outcomes.

Theory of change vs. program logic

Theory of change: an explicit theory or rationale for how and why a program will achieve intended outcomes.[81] It articulates the hypothesised causal relationships between a program's activities and its intended outcomes and identifies how and why changes are expected to occur, by providing evidence and theory. In doing so, the theory of change comprises a change model (the changes the program intends to achieve) and an action model (the activities that will lead to those changes). A theory of change must be plausible, doable and testable.

Program logic: a visual representation of how a program will achieve its goals, including short-, medium- and long-term outcomes. It comprises a detailed representation of inputs, activities, outputs, outcomes and impact. Program logics are a more simplistic representation that graphically sets out the different elements of a program and establishes links between these different elements.

4

UNDERSTAND WHAT TO MEASURE AND EVALUATE

You now have a clear picture of what your program is trying to achieve. It is time to think about what you will measure and evaluate, and how.

WHAT IS EVALUATION

Evaluation **is a systematic process of understanding the merit, worth or significance of a program that aims to achieve social change by combining evidence and values.**[82] Ultimately evaluation seeks evidence to answer the questions such as has this made a difference? Is this worth doing?

EVALUATION QUESTIONS

Evaluation is an activity that should directly inform decision making. It is important to think about what you want to know from the evaluation (i.e. your 'key evaluation questions') as these will determine the appropriate evaluation type. What decisions do you need to make or are you likely to make based on the evaluation findings, or what do you need to know to help staff and other stakeholders deliver a better program? This thinking is often supported by developing evaluation questions.

The Organisation for Economic Cooperation and Development (OECD)[83] developed six evaluation criteria that can also serve as guidelines or a thinking tool for selecting evaluation questions:

- **Relevance:** Is the program doing the right things? The extent to which the objectives and design respond to beneficiaries', global, country, and partner/institution needs, policies, and priorities, and continue to do so if circumstances change.
- **Coherence:** How well does the program fit? The compatibility of the program with other initiatives in a country, sector or institution.
- **Effectiveness:** Is the program achieving its objectives or outcomes?[viii] The extent to which the program achieves, or is expected to achieve, its objectives, and its outcomes, including any differential results across groups.
- **Efficiency:** How well are resources being used? The extent to which the program delivers, or is likely to deliver, results in an economic and timely way.
- **Impact:** What difference does the program make? The extent to which the program has generated or is expected to generate significant positive or negative, intended or unintended, higher-level effects.
- **Sustainability:** Will the benefits last? The extent to which the net benefits of the program continue or are likely to continue.

It may be that not all these criteria are relevant to your program or decision-making environment, but they can guide you into the questions you would like your evaluation to answer.

viii Note that outcome evaluation is a type of evaluation that assesses program effectiveness. Outcomes measurement is a process that is part of assessing a program against this evaluation criterion.

Measuring everything you want to may not be feasible due to a range of constraints (resources, time, access to stakeholders). This is a good time to prioritise what to measure (in particular, outcomes). You need to consider your evaluation questions – those are the questions that you want answered. Think again about the following:

- Your purpose and aims (what are you evaluating and why).
- Key stakeholders (whose outcomes will you measure?).
- Time (what is your timeline for data collection?).
- Skills (do you have staff to collect quantitative and qualitative data).
- Funding (can you afford it?).

Keep in mind that....

'Not everything that counts can be counted and not everything that can be counted counts'[ix]

Examples of evaluation questions for Sport

To what extent was the program appropriate for the needs of school students?

- Was the program effective in motivating students to be active? If so, what changes in activity levels did students experience?
- Were there particular cohorts of students (gender, culture, disability, etc) that were more responsive to the program? What were the barriers in reaching other students?
- What factors in the school environment supported program participation?
- What could be improved?

TYPES OF EVALUATION

Evaluation can occur throughout the lifecycle of a program. A **formative evaluation** describes evaluation activities that take place during a program's implementation, with the aim of improving its design and performance. A **summative evaluation** will occur at the end of a program, or when the program is well-established enough to be able to make assessments on whether the program has achieved its intended goals.

'When the cook tastes the soup, that's formative evaluation; when the guest tastes it, that's summative evaluation'.[84]

Depending on its purpose, your evaluation may be:[x]

- **Process evaluation:** investigates how a program was established and implemented or delivered.
- **Outcomes evaluation:** explores the changes occurring as a result of a program.
- **Impact evaluation:** examines the long-term, sustained changes as a result of a program. As noted previously the term impact evaluation is also used to encompass the evaluation of both short and medium-term outcomes but long-term impacts.
- **Economic evaluation:** studies the benefits and costs of a program, including the cost offsets of a program, and the value generated as a result.

Process and outcome evaluation are two common types of evaluation (basic steps for these are shown in Figure 10 on next page). They often go together.

ix Quote linked to the work of sociologist William Bruce Cameron.
x Note there are many different types of evaluation. Those mentioned throughout this document are just a few of these.

Process evaluation examines *how* a program has worked and explores details about different elements of the program, the mechanisms that connect them, contexts and actors that have driven observed outcomes and types of activities undertaken. Process evaluation may be either formative or summative, depending on whether its purpose is to inform decisions about implementation (formative) or to inform decisions about investment or whether or not to continue a program (summative). It can be useful to link a process evaluation to a program's theory of change or logic model, as theories of change can show which elements are important to include in data collection and analysis. Theories of change and logic models can also be refined through process evaluations as this type of evaluation identifies the most important activities, and how their implementation should be working in practice.[85]

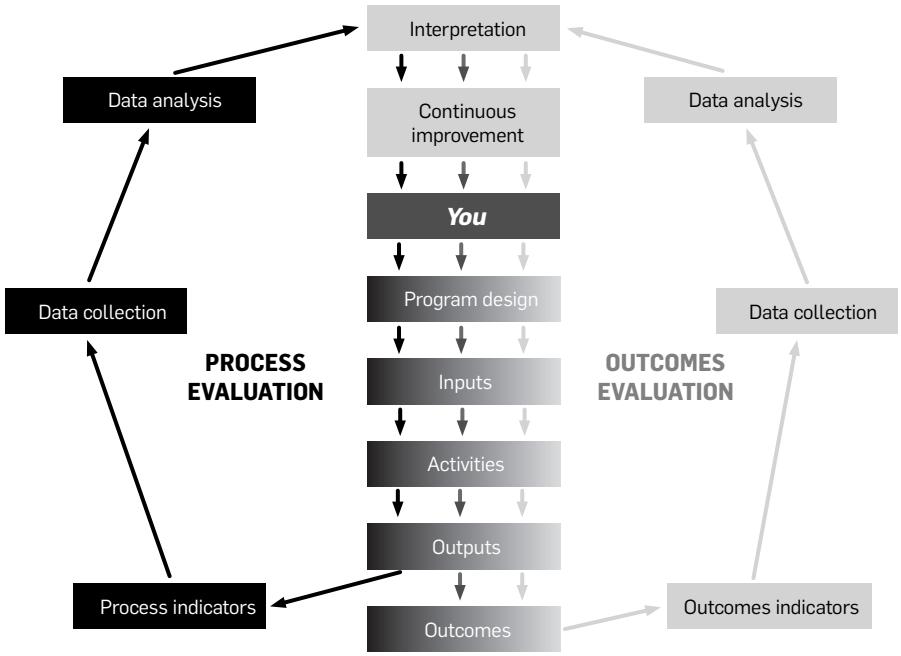
Outcome evaluation examines the changes that occurred as a result of a program. It often examines 'change' in terms of knowledge, skills, attitudes and behaviour change in the population targeted. While there are merits and limitations associated with both approaches, they provide good options to isolate and evidence the outcomes of a program.

How process and outcome evaluations work together

Process evaluation may explain why certain outcomes were or were not achieved. It helps to identify if some outcomes were not achieved due to program failure (i.e. has the program failed to achieve a set of outcomes for its beneficiaries?) or implementation failure (i.e. the program was not implemented as intended, hence the outcomes could not have been achieved). An example of implementation failure in the *Sport* case study would be if the information sessions and resource packages to parents were not delivered twice a year (i.e. did not deliver the intended activities).

Figure 10:

Diagram showing how process and outcome evaluations work together



The 'Counterfactual'

A 'counterfactual' is an evaluative method of understanding what would have happened *without* a program by comparing observed results or outcomes between groups that have and have not been affected by that program.

Impact evaluation is the assessment of the extent to which long-term, sustained changes resulted from the program activities. This type of evaluation is more likely to influence policy. It can be conducted at some point throughout the delivery of the program (for ongoing programs) when according to the theory of change the impact would have been achieved at least for a group of program participants. The key element of impact evaluation is the 'counterfactual', or **what would have happened had the program not been implemented**. Being able to compare the 'do nothing' scenario with the outcomes achieved from the program will provide evidence for the changes produced by the program.

In a **Randomised Control Trial** (RCT) design, eligible (and consenting) potential participants in a program are randomly assigned to the program and 'treatment as usual'. Outcomes measurement proceeds at baseline (i.e., before randomisation), during the period of the program, at exit and beyond for both in the program and the treatment as usual group and then an assessment is made of the difference in outcomes and impacts for the two groups. Proponents of the RCT design argue that the RCT design is the best (and possibly only way) to robustly estimate the differential social impact of a program. However, in the social context the RCT design is almost impossible to implement in a double-blind form (as is the standard in the medical clinical case), requires significant resourcing, can face perceived ethical concerns from organisations and their staff, and resistance from potential participants.

The RCT impact evaluation design is often referred to as an experimental design. In a **quasi-experimental design** (QED), there is no randomisation of potential participants in the program but both participants and those who could have been participants are followed throughout and an assessment of the differences in outcomes and impacts for the two groups. Unlike the case of the RCT there is no presumption that the group of program participants and the group of non-participants are equivalent in background and composition. As such, the QED requires robust statistical modelling of outcomes controlling for measured characteristics of members of the two groups. As with the RCT design the QED design is costly to implement and requires a high level of sophistication on the part of the evaluation team.

If an RCT and QED design is not possible to implement for various reasons including cost, organisations are still encouraged to build a comparison reference that will allow for reasonable estimates of the differential impact of a program. This comparison reference can be built from understanding the history of the participant group, an examination from secondary data sources of the trajectories of those who could well have been participants in the program, and an assessment of the potential impact of external confounding factors on outcomes. This design is the **Single Group with comparison** design and is one that is accessible to social purpose organisations.

Table 3 below shows examples of outcome measurement and evaluation approaches that aim to measure change, and their data collection frequency. Several of these approaches rely on measuring change compared to what would have happened had the program not been implemented (the 'counterfactual'). The term impact evaluation is sometimes reserved for those designs that establish a counterfactual to show long term change.

Impact evaluations are not cheap or straightforward. There are also ethical considerations relating to using a control group and whether they are culturally safe or appropriate. There are also critiques about whether isolating causal factors is possible in complex environments.

Turn to Appendix 4 to learn about techniques used to establish comparison groups, such as RCTs and QEDs.

Economic evaluation focuses on how to best allocate resources (human time and skills, equipment, etc) towards programs. If we have unlimited resources, we only need to know the efficacy or effectiveness of a program – and then we can simply choose the most effective option. However, in the real-world, resources are limited and we need to know "Where can we get the most bang for our buck?" Economic evaluations help answer this question through comparing the costs and outcomes of different program options. Economic evaluations help to make decisions to optimize use of resources, decide between different program options and demonstrate the benefits of different programs in monetary terms. Types of economic evaluations include Cost-Benefit Analysis (CBA), Cost-Effectiveness Analysis (CEA), Cost-Utility Analysis (CUA), Social Return on Investment (SROI) and Cost Minimisation Analysis (CMA).

A major focus in economic evaluations of social programs is on the estimation of **cost offsets**. A cost offset is the reduction in costs that accrue in non-program areas as a result of outcomes achieved in the program. For example, a homelessness program achieves high rates of stable housing which in turn results in lower emergency use of hospitals. The reduction in hospital costs is a cost offset to the cost of the homelessness program itself.

Table 3:

Examples of outcomes measurement and evaluation designs and data collection frequency

Stakeholder groups	Examples of role description and contribution to outcomes measurement
Measurers (e.g. practitioners; researchers; evaluators; consultants)	Design and implement outcomes measurement and evaluation frameworks, analyse and present the results
Program managers and funders (e.g. government; corporates; NGOs; enterprises; business)	Possible funders of outcomes measurement and evaluation; inform design; provide data if relevant; use evidence from outcomes measurement and evaluation to write policy; to inform future investment; to develop or refine initiatives
Beneficiaries (clients/consumers) and communities	May be partners in the design and implementation of outcomes measurement and evaluation; primary providers of knowledge and data
Innovators & implementers (e.g. social purpose organisations; communities)	Inform design (may be conducting or partners in the design); data collectors; data providers and reporters
Others (e.g. partner organisations; service providers; advocacy groups; family members; etc)	Data providers; may inform the design

Case study: 'Beyond the bottom line' – Economic evaluation

The "Beyond the Bottom Line" study[86] evaluated the impact of social enterprises offering open employment opportunities for people with disabilities. It focused on the effectiveness of the Social Return on Investment (SROI) methodology in this context. The study applied an adapted form of cost-benefit analysis using SROI, which enables the inclusion and valuation of outcomes typically omitted from traditional forms of economic evaluations. The SROI forecasting process in this study utilised two models: one focused on calculating the financial outcomes of the social enterprise, and the other incorporated both financial and social outcomes.

Data was collected from various stakeholders, including people with disabilities, their families, staff, social investors, and others (n = 17). The study also included expert input and a sensitivity analysis to assess the robustness of the results. Findings showed a social return of $1.47 to $2.65 for every $1 invested over three years, and $8.48 to $12.63 over 20 years. These findings offer useful insights for the transitioning ADE sector, social investors, and policymakers seeking reliable impact metrics.

The use of the two SROI models was crucial for highlighting the added value of social benefits over and above the financial impacts of the social enterprise. The findings emphasised the long-term impact of such an intervention not only on the long-term social and economic trajectory of people with intellectual disabilities over the course of their working lives but the wider societal benefit in terms of reduced welfare and increased taxation. On an organisational level, the SROI methodology proved valuable in effectively communicating the potential benefits to investors, while also strengthening the organisation's mission to provide integrated, long-term employment opportunities for individuals with disabilities of all ages.

'NEXT-GENERATION' EVALUATION APPROACHES

For many decades the main focus of evaluation was program evaluation. Programs typically have well-defined boundaries. It is no surprise that traditional evaluation tools and methods tend to be based on the idea that change is incremental, logical and linear[87], and that activities lead to outcomes in a straightforward way that can be tracked. **There is now more acknowledgement of the increased complexity of social issues** – both in terms of the environment in which they are delivered, the complexities of effectively addressing those issues and the nature of programs that attempt to address them[88],[89]. For instance, many initiatives we are evaluating now involve collaborations or coalitions and seek to address strategy, policy, systems change or multi-factor causal factors such as the social determinants of mental health.

'Next-generation' approaches to evaluation have emerged in response to this complexity and rather than the focus being on making summative claims, they seek a **greater focus on learning and adaptation**. They also position the evaluator in terms of their role as facilitator and coach, rather than an external 'objective' source of knowledge (refer to Table 4 on next page, sourced from Zappala, 2020[90]). Next generation evaluation approaches should be seen as complementary to traditional evaluation approaches and methods rather replacing them.

Table 4:

Differences in focus between traditional and next generation evaluations

Traditional Evaluations	Next-Generation Evaluations
Accountability	Learning
Compliance	Capacity building
Formative/summative	Developmental
Logic models	Systems approach
Accounting for impact	Demonstrate learning/adaptation to change
Evaluator as external/objective	Evaluator as facilitator/coach
Pre-determined goals	Shifting goals/adaptation
Fixed plan	Variable/emergent plan
Anticipated outcomes	Unintended outcomes
Individual outcomes	Holistic outcomes

Case study: Community Connections Program - Developmental evaluation

A developmental evaluation of the Community Connections Program (CCP) was conducted by CSI Flinders University in 2022.[91] CCP is managed by the Department of Human Services (DHS) South Australia. To implement CCP, DHS funded a range of government and non-government partner organisations across the state to deliver short-term programs designed to help individuals build connections with communities, social networks, and services to increase their independence. Given that CCP was a new program, it was expected to evolve over time to better address emerging needs of participants and partners. To support this dynamic process, CSI conducted a developmental evaluation.

The evaluation primarily focused on the early stages of CCP nine months into its implementation, with the goal of supporting development and assessing its effectiveness to guide improvements and planning beyond June 2023. Data included a quantitative survey data from DHS' Results Reporting Data Dock, qualitative data from partner quarterly reports, and focus groups with both partners and participants.

The developmental evaluation played a key role in assessing the early impact of CCP and identifying key areas for improvement. The evaluation revealed the program's positive impact on participants, particularly in enhancing social connectedness and quality of life. Partners also highlighted CCP as a pioneering initiative that promoted collaboration over competition, with strong examples of partner organisations working as a system through referrals and sharing resources. The evaluation uncovered areas for improvement such as data collection and management, partnership structure, program flexibility and the management of cases with varying levels of complexity. Based on these findings, recommendations for program improvements were developed and provided to DHS, resulting in substantial changes to the program structure in 2023. This demonstrates the relevance of developmental evaluation in providing real-time feedback that informs ongoing program refinement and ensures it continues to meet the evolving needs of participants and stakeholders.

Two of the more common examples of next generation evaluation approaches are developmental evaluation and realist evaluation.

Developmental evaluation assists social innovators develop social change initiatives in complex or uncertain environments.[92] It involves asking evaluative questions and gathering information to provide feedback and support decision-making along an emergent path in the context of innovative initiatives and development.[93] It primarily focuses on adaptive learning (rather than accountability) by providing real-time feedback and generating learning.[94] The evaluator is considered a member of the program delivery team, whereby they actively support decision-making and surface innovative strategies and ideas through data collection and analysis. Developmental evaluations take a flexible approach and enable evolution of measures and monitoring mechanisms as understanding of the situation deepens.[95,96]

Realist evaluation seeks to understand how and why programs achieve different outcomes for different people using a context, mechanism and outcomes model.[97] The key to a realist evaluation is clearly defining how context, mechanisms and outcomes are operationalised.[98] One of the common confusions is in defining the difference between the intervention and mechanisms.[99] Typically in a realist evaluation, the intervention relates to something provided within an organisation, such as an education program or a physical activity program.[100]

Other examples of next-generation evaluation approaches include:

- Systems- and complexity-informed evaluation.[101]
- Blue marble evaluation.[102]
- Meaningful evaluation.[103]

A range of questions that each type of evaluation may answer is provided in Figure 11 on next page.

Figure 11:

Example evaluation questions for different types of evaluation

PROCESS EVALUATION

- Was the program implemented as intended?
- What enabled or constrained its implementation?
- Did the program reach those it was intended for?
- Was the program suitable to the needs of those it was intended for?
- What were the strengths and opportunities in how the program was delivered?
- Are there opportunities to improve data collected through the program?

OUTCOMES / IMPACT EVALUATION

- What was achieved - and for whom?
- What worked best - and why?
- What outcomes were achieved and how reliably?
- Are there differences in outcomes between client groups?
- What was the broader (social) benefit or impact?
- Was this role appropriate to the proponent?

REALIST EVALUATION

- For whom will this program work and not work, and why?
- In what contexts will this program work and not work, and why?
- What are the main mechanisms by which we expect this program to work?
- If this program works, what outcomes will we see?

DEVELOPMENTAL EVALUATION

- What is emerging as the innovation takes shape?
- What do initial results reveal about expected progress?
- What variations in effects are we seeing?
- How have different values, perspectives and relationships influenced the innovation and its outcomes?
- How is the larger system or environment responding to the innovation?

ECONOMIC EVALUATION

- What was the ratio of costs to benefits?
- What is the most cost-effective option?
- Has the program been cost-effective (compared to alternatives)?
- Do the outcomes of the program represent value-for-money?
- Is the program the best use of resources?

SYSTEMS- AND COMPLEXITY-INFORMED EVALUATION

- What are the sociocultural, environmental and economic factors that contribute to the problem at hand?
- How do they intersect with this program?
- What are the systems levers — i.e. policy options, stakeholders and assets — that can change the situation?
- What would be the broader consequences of influencing these system levers?

Case study: Connecting Pathways to Employment - Realist evaluation

A realist evaluation of four work integration social enterprises (WISE)[104] explored how the organisation contributed to various health and wellbeing outcomes of young people from disadvantaged backgrounds. The evaluation aimed to uncover the complex and dynamic interactions between processes and policies, spaces, relationships, and organisational and personal factors that influence the health outcomes of these young people. The research used a range of methods, including semi-structured interviews, participant observation, and engagement workshops.

In this evaluation of how a social enterprise as a whole was impacting on wellbeing a different use of context, mechanisms and outcomes was used. The **organisation** itself was treated as the intervention. This is due to the purposeful design of the social enterprise model which, through its structures and processes, seeks to be inclusive and health-promoting. The **context** was conceived as the particular backgrounds and health challenges of the employees of the organisations (micro context) and the broader social and economic environment (inclusive of industry norms and government policy context). The **mechanisms** or processes of change were conceived as the psychosocial pathways by which participants experience the organisation and the outcomes were the self-reported changes to their health and well-being.

Thematic analysis was conducted to identify patterns in how participants described their experiences and topics that were important to them. A realist evaluation approach was used to refine these themes, revealing how the social enterprises provided a sense of control over work, reduced stigma about sharing mental health issues, and fostered inclusion and belonging for people from different backgrounds. The findings emphasized the heterogeneity and complexity of the mechanisms and health outcomes experienced by participants, depending on their personal and work situation. The WISEs were found to offer different structures and supports, addressing various health situations and backgrounds, which varied within and between each of the WISE cases.

CONTINUOUS QUALITY IMPROVEMENT

Some organisations seek to deliver social impact through the culture and structure of their organisation. It is not through delivering programs per se that they seek a social impact rather through the operation itself of the organisation. When settings such as workplaces and schools strive to ensure that their culture and operations are always achieving social benefit then a program approach to evaluation is not suitable. In these situations, a continuous quality improvement (CQI) method is required such as annual surveys or some other data method. The purpose is to continually monitor that the organisation is performing well against the social indicators of importance. An advantage of the CQI method is being able to harness the Hawthorne effect as an intervention technique to motivate improvements in performance and attainment of goals.[105],[106] The Hawthorne effect is a classic threat to validity where change occurs not through the intervention but due to the motivation of knowing that you are being observed.[107] However, from a CQI perspective this is not a threat to validity but an implementation opportunity,[108] as data itself can be a potential leverage point to positively influence performance and attainment of goals.[109,110]

POLICY CHANGE AND OUTCOMES MEASUREMENT AND EVALUATION

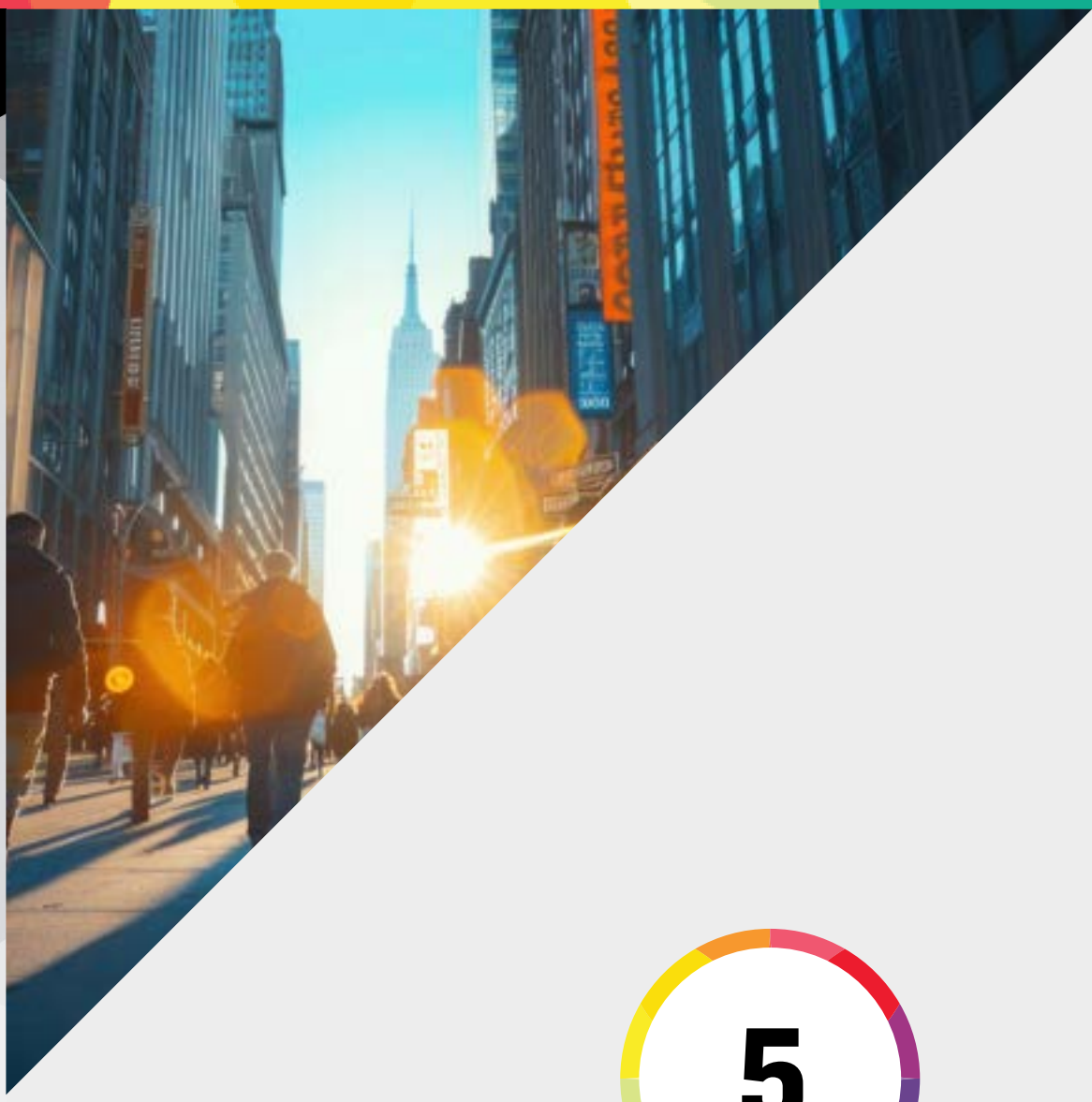
For many social issues such as family violence, homelessness, and employment, policy and/or legislative change is required to achieve social impact. Outcomes measurement and evaluation for policy change can take two different directions. Broadly speaking, the first is a technical approach to evaluating the outcomes of a specific policy as it is being implemented, much in the same way as evaluating the impact of a specific program. The second approach is evaluating the actual process itself of how the policy change occurred.[111] This could involve an analysis of the roles of institutions, how different ideas and beliefs shaped the policy process, the roles of different policy actors, and the political and contextual factors involved in the timing of any change.[112] Evaluation methods to understand the process of policy change can include interviews, focus groups, and social network analysis.

Case study: Policy change for upscaling a respectful relationships program

An evaluation was conducted on how the Respectful Relationships in Schools Program in Australia[113] was upscaled so quickly after the pilot phase of the program was delivered across the State of Victoria. This program supports schools and early childhood settings to promote and model respect, positive attitudes and behaviours.

The evaluation involved interviews with project staff, community-based organizations, and Department of Education and Training (DET) representatives to understand their perceptions, challenges, and strategies for upscaling the program. The major policy change to addressing respectful relationships in schools occurred as a result of a combination of factors. Firstly, a high profile personal tragedy propelled family violence into the national spotlight. Simultaneously, the organization leading the Respectful Relationships program leveraged their strong networks to position the program within the education department, enabling the government to quickly respond to the issue. Additionally, having a program available that could be scaled quickly contributed to timely policy change.

The study emphasizes that political will is not a stand-alone factor as depicted by up-scaling models, but rather is the end point of a complex process that involves many elements, including the establishment of networks and aligned programs that can capitalize when opportunities arise.



5

DEVELOP AN OUTCOMES MEASUREMENT AND EVALUATION FRAMEWORK

Your evaluation questions give you an indication of what you wish to learn from the evaluation and outcomes measurement work. The next step is to develop a framework that clearly outlines, step by step, how you will define and collect the evidence needed to answer your evaluation questions and conduct additional supporting activities (such as following ethical principles). An outcomes measurement and evaluation framework will be your guiding document for mapping the process and undertaking this work.

ESTABLISHING FOUNDATIONAL FRAMEWORKS

There are many ways to create plans for outcomes measurement and evaluation work. There are also numerous terms such as 'evaluation plan', 'evaluation and measurement framework', 'outcomes measurement framework', 'monitoring and evaluation plan'. We present here some guidance for a few of the different terms for commonly used frameworks. However, we recognise that it is difficult (and often not necessary) to find fixed definitions for these terms, because language in this space is often fluid and flexibly applied (our advice: **don't get too caught up in the wording**).

An **evaluation framework** is an overarching document that presents all the elements that may be relevant to how you are planning to systematically assess the merit, worth or significance of your program and its components. It may include, for example: the purpose of your evaluation and outcome measurement work, your evaluation questions, the approach and the type/s of evaluations you have chosen, the evaluation design adopted (e.g., RCT, quasi experimental, single group with comparison), methodology, governance arrangements, how you will meet ethical requirements, a program logic and theory of change and an outcomes measurement framework. It may also include information about the analytical frameworks you will apply to your data and how you will translate the knowledge you have generated into action (see below).

An **outcomes framework** (also referred to as an 'outcomes hierarchy') outlines, conceptually, the outcomes that are expected, relevant and in the scope of influence for your program. This is a good starting point for developing an **outcomes measurement framework.** This document will bring together the **outcomes** that you've identified and intend to measure, the **indicators** or measures for the outcomes, the **data sources** you will use to quantify or qualify those indicators, and the **timing for data collection**. Your outcomes measurement framework is often a component of a broader evaluation framework (or it can be a stand-alone document that is regularly updated). You will learn how to select indicators in this section and section 6 discusses the various types of data you can collect to quantify and qualify indicators. An **outcomes measurement and evaluation framework** could be an all-encompassing document that brings together all your thinking!

WHAT TO INCLUDE IN AN OUTCOMES MEASUREMENT AND EVALUATION FRAMEWORK

An **outcomes measurement and evaluation framework should include an overview of the steps you plan to take** (which may align to the process described within the Roadmap). It may include:

1. Purpose, aim and objectives of the evaluation
2. Overview of the type of evaluation being conducted
3. Key evaluation questions and criteria
4. Methodology and limitations
5. Governance arrangements and ethics
6. Program logic and theory of change (if it has been developed)
7. Outcomes measurement framework including indicators and data sources
8. Data collection plan
9. Analytic methods or frameworks
10. Plan for knowledge translation

Your outcomes measurement and evaluation framework should be developed in consultation with relevant key stakeholders such as people with lived experience of the issue at hand.

INDICATORS

Indicators are the measurable markers that can be used to answer evaluation questions and to measure success in achieving outcomes. They are often extremely precise, but based on a rationale that if one thing is going in a positive direction, it is likely that other things are also tracking in that direction. For instance, if school attendance rates for at-risk students are increasing, one might assume that this could also be an indicator (or a proxy indicator) that student wellbeing might be improving and/or literacy and numeracy outcomes might also improve. Other data sources alongside this indicator can confirm if these assumptions are sound.

Indicators can be expressed as percentages, proportions, numbers, ratios, perceptions, behaviours, satisfaction or quality. **They can be expressed quantitatively or qualitatively.** Indicators can be a single measure capturing a standard or condition at a certain point in time, such as the qualitative perceptions of key stakeholders or the proportion of participants living with a mental health condition. Indicators may also be a composite made up of several measures, such as the Kessler psychological distress scale,[114] which measures ten aspects of psychological distress but reports this as one value between 1 and 50.

The language of indicators

Indicators capture at one point in time participation rates, individual behaviours, incidence, prevalence, and attitudes and other measures. They can be formulated as a proportion of respondents reporting x, level of x, satisfaction with x, knowledge of x, awareness of x, attitude towards x, skills, level of confidence.

Examples include:

- Perceptions of program relevance.
- Satisfaction with own health (on a 0 to 10 scale).
- Number of hospital admissions in the past 6 months.
- Proportion of respondents able to identify health risks.
- Number of drinks per day
- Number of young people with Cert II qualifications.
- Proportion of children enrolled in early education.
- Perceptions of sustainability of benefits.

They can also be established measurement scales, like the Rosenberg self-esteem scale (measure of self-esteem), the Personal Well-being Index, or mental health scales (like Kessler 10, Kessler 6, or DASS21).

STEPS TO DEVELOP INDICATORS

1. Allow for time and resources to review indicators. Consider the scope of the evaluation, how long do you have, can you engage stakeholders appropriately, and do you have resources in place?
2. Search for existing indicators used by industry, academic, government, national and international sources, and national and international indicator banks (see Appendix 2 for potential sources of indicators). Drawing on existing indicators will often ensure your indicators respect all technical criteria. Make sure your indicators consider equity (i.e. who is benefiting the most and least from the program and its evaluation).[115]
3. Review the technical and contextual suitability of indicators, noting that there may not always be validated or standardised measurement tools that are appropriate for your program. This will be team-work - engaging stakeholders helps to understand if indicators are appropriate and acceptable.
4. Select indicators with consideration as to whether some were prioritised by stakeholders and whether gaps were identified (i.e. need to develop new indicators).
5. Consider new indicators if your existing indicators are not a good fit for your program.
6. Choose only those indicators that are useful, not only those that can be measured.

TYPES OF INDICATORS

Within your framework, you can now define your outcomes and indicators further. You can do this by looking at the logic model you developed, your stakeholder groups, your evaluation questions and flagging which outcomes you need to measure for the type of evaluation you want to conduct.

Various criteria for outcome indicator selection have been developed over the past two decades, including SMART (Figure 12 below).[116]

Figure 12:
SMART indicators

You can also differentiate further between technical and contextual indicators for measuring outcomes.[117] Technical criteria refer to the extent to which the indicator is a good measure for your outcomes. For example, whether the outcome indicator is validated (is there evidence to support that the indicator measures what it intends to measure?), or reliable (does the indicator produce consistent results over time?). Contextual criteria look at surrounding characteristics that can help you decide whether the indicator is a good fit for the outcome, given your program context. For example, is the indicator acceptable (will the participants be comfortable to answer certain questions?) or is it feasible (is it practical to collect the respective data?). See Appendix 3 for example templates to measure these criteria.

Regardless of our efforts to select or develop 'good' indicators, as the terminology suggests, an indicator is *indicative* of the outcome it seeks to measure. Two or more indicators may be needed to measure an evaluation question or outcome. For example, improved youth mental health can be measured through the proportion of young people reporting a mental illness in the past 12 months but also through the Kessler score (K6 for youth). The two indicators capture the frequency and intensity of mental illness in youth, therefore both are needed to assess change.

It is a good idea to ensure you have quantitative and qualitative indicators for your evaluation criteria. Indicators often imply quantifiable concepts measuring how much/many/often. However, it is important that they also capture qualitative responses, such as perceptions, attitudes and feelings (e.g. changes in quality of life; feelings of anxiety). Qualitative data is also useful for hearing from people in their own words, which may be especially useful when measuring the impact of programs for people who may not respond well to structured questions or speak English as an additional language.

See Table 5 below for some examples of quantitative and qualitative measures for *Sport* and the next section (Section 6) for insights into data collection techniques for quantitative and qualitative indicators formulated at individual, community or societal level.

Table 5:
Outcomes and indicators - fictional case study, *Sport* program

Outcome	Indicator
Students engage in regular exercise (medium-term)	Proportion of students participating in 2-3 after school sport activities per week Proportion of children undertaking at least 60 min of physical activity each day Feelings of inclusion (qualitative)
Students are physically healthy (long-term)	Proportion of students with healthy BMI Proportion of students with reoccurring health conditions
Students lead happy lives (medium/long-term)	Quantitative indicator: • All things considered, how happy are you these days? Please indicate on a 0 to 10 scale, with 0 being least happy and 10 being most happy. Qualitative measure (questions in interview): • Please tell me how you feel during a normal day. • What is it like waking up early and going to school? • How do you feel about being in school? • Do you have a favourite part of the day that makes you feel happy? What do you do then? Why is it your favourite? Why do you think it makes you feel happy?

Case study: Community Services Outcomes Tree

The Community Services Outcomes Tree (https://communityservicesoutcomestree.com) was developed by the CSI at Swinburne University of Technology, in partnership with Uniting Vic Tas.[118] The Community Services Outcomes Tree is designed to help community sector organisations measure the effect they are having on individuals/families' lives. Measuring these outcomes can help to guide service delivery, support advocacy and secure funding. The framework encourages a 'whole of life' approach and recognises the way in which life domains interrelate. For example, while an organisation may focus on education it would be worthwhile to consider the impact of the service in other areas of a person's life such as employment, or social inclusion etc.

The outcomes framework is presented in the form of a 'tree' with twelve leaves. Each leaf is an outcome domain, for example 'housing'. Each domain is comprised of a set of outcome areas. These represent the key outcome topics or themes found in the literature, in other outcomes frameworks and measurement instruments. This framework articulates, for the first time, the range of outcome areas that the community services sector aims to contribute to.

The website also offers guidance on outcomes measurement as well as a tested survey approach. For each outcome area there are two key questions that can be used:

Question 1: seeks to measure the amount of change that has occurred for the service user in relation to the outcome area.

Question 2: seeks to measure the level of contribution the community service has made to this change.

Other suggested survey questions include the barriers that hindered the attainment of outcomes, and areas for service improvement.

Case study: Together Home Evaluation

The Together Home evaluation[119] highlights the importance of carefully identifying and prioritising outcomes to achieve meaningful impact. Together Home, a Housing First program, aimed to transition people experiencing rough sleeping into long-term housing. CSI UNSW's evaluation focused on identifying critical program outcomes, ensuring alignment with the Housing First model.

Key outcomes prioritised and measured included:

- **Housing retention:** A retention rate of 85.8%, demonstrating the program's success in providing stable housing.
- **Health improvements:** Clients reported better health and reduced hospital admissions, with hospital use dropping from 35% to 20%.
- **Wellbeing enhancement:** Wellbeing scores on the Personal Wellbeing Index (PWI) rose from 5.4 to 8.3 after 18 months.

These outcomes were prioritised based on their alignment with the program's overarching goals and their relevance to stakeholders, including policymakers and service providers. The findings informed recommendations for increasing funding for permanent housing solutions, integrating mental health services, and tailoring support to client needs.

This project demonstrates how identifying and prioritising outcomes can guide impactful evaluations and deliver actionable insights for program improvement and policy development.

Tips for developing your outcomes measurement framework

Once again, use flip chart paper and post-it notes (or if working online, a document that can be shared and edited by all participants. You may want to have one of the participants as a scribe that leads the note-taking while the rest of the group brainstorms). Invite your team and key stakeholders (including people with lived experience) to participate in the following process if possible.

1. Split the butcher's paper (or the document you work on) into four columns: outcomes, indicators, data sources, target population and timing for data collection. Write an outcome on a post-it note and place it in the outcomes column.
2. Move to the next column and add the indicator for this outcome (on a separate post-it note).
3. Continue with data source (question to include in a survey, administrative, interview, etc) and the target population and timing for collection (e.g. young people, pre-program participation and 6-months into the program).

You might find that outcomes are organised by short-, medium- and long-term or by for the main beneficiaries, then other stakeholders. Make sure to discuss your logic model and evaluation questions to agree on which outcomes should be measured. This will provide you with insights from a range of people and agreement over outcomes and indicators, as well as data sources and timing for collection.

LINKING OUTCOMES AND INDICATORS TO ESTABLISHED FRAMEWORKS

Where appropriate, it is good practice to link outcomes and indicators within your outcomes measurement and evaluation framework to those of established local, state, national and international frameworks. This ensures that the outcomes you are working towards, and data that you are collecting to measure progress towards them, align with outcomes that have already been identified as important. For example, you can look at published frameworks, strategic plans and other policy documents of the following to see whether outcomes and indicators have been identified at different levels:

- **Local** councils or leadership/advocacy groups.
- **State, territory or national** governments, collective impact groups and peak bodies, such as the 'Measuring What Matters' framework[120] developed by the Australian Government and the Nest Wellbeing Framework[121] developed by ARACY.
- **International** frameworks and organisations such as the Sustainable Development Goals[122] developed by the United Nations.

Case study: The Nest

The Nest[123] is an evidence-based framework for national child and youth wellbeing. It was developed by the Australian Research Alliance for Children and Youth (ARACY) in 2013, through extensive consultation with children, families and experts. The framework outlines six critical and interconnected dimensions of wellbeing that are key for children and young people to thrive (Figure 13 below).

Figure 13:
The Nest framework

The Nest can be applied in several different ways such as aligning program outcomes and indicators for direct service delivery for families, policy and strategy development, place-based work and monitoring and evaluation.

Good practice tip

It is good practice to collate the program logic and outcomes measurement and evaluation framework into a single data collection plan. This will facilitate your understanding of the program from inputs to outcomes, the data collection needs and potential approaches to assess outcomes and impact, as well as associated risks and assumptions. Turn to Appendix 3 for data collection plan templates.

6

COLLECT DATA

To complete your outcomes measurement and evaluation framework, **you must decide the most appropriate data collection tools to quantify or qualify the indicators that you identified** through the previous step. See Appendices 3 and 4 for examples of how to identify appropriate data sources to answer key evaluation questions and measure outcomes. Here we explore different ways to collect and monitor data.

Both quantitative and qualitative data can and should be collected for evaluation and outcomes measurement. It is important to collect data at intervals relevant to the scope of the evaluation (e.g. pre-program, throughout the program, half-way through the program, end of the program and/or a few weeks/months after) to assess change in indicators and the extent to which outcomes are achieved.

Baseline data and benchmarking

Data collected prior to the program is called **baseline data**. This data can help you compare program participants to the general population (e.g. by comparing with national statistics). It also serves as the reference point, helping you make conclusions about the change by comparing how an indicator has changed as the program progressed. Think about whether this baseline data is readily available for your program, or how you could collect it. For example, such baseline data could be available in administrative records.

In addition, data from secondary sources such as population data, or from the evaluation of programs similar to yours can serve for benchmarking. **Benchmarking** investigates how the target population compares to larger populations, or the extent to which outcomes were achieved, compared to other programs.

QUANTITATIVE METHODS

Surveys, administrative data and secondary data are quantitative data sources most frequently used in outcomes measurement and evaluation.

Surveys

Surveys are data collection instruments that may be standardized and are usually administered face-to-face, online or by phone to generate quantitative data. They may also collect qualitative data, often regarding people's experiences and attitudes.[xi] Surveys can be an efficient way of collecting data as they reach large numbers of people for a relatively low cost and can be repeated to track behaviour changes. Response rates, however, can be low, which can jeopardise the validity of the data collected.

Surveys can be administered at program, organisation, sector, or national level. When deciding what type of survey to administer consider:

- Your target population (e.g. are they more likely to respond online or face-to-face? Consider their demographics, literacy, computer literacy, access to technology and likelihood of responding).
- Budget (online surveys are cheaper to administer than phone or face-to-face).
- Type of questions (some questions might need visual supports; complex questions may be easier to design in online formats or to explain face-to-face).
- Will respondents be more likely to share accurate information if the interviewer is present or absent? Does the demographic (e.g. gender or ethnicity) of the interviewer matter?

xi While survey questionnaires are structured and offer options for answers (such as agreement scales), it is common to also include open-end questions, where respondents have the opportunity to offer additional information.

Program administrative data

The term administrative data is used in outcomes measurement and evaluation both to refer to program data collected for all participants in a program and data held in public datasets of program participants.

In the case of program administrative data, a case worker, for example, might record information about a client after each encounter. Headline data that an organisation uses in annual reporting (e.g. proportion of female clients) is also administrative data. While the primary use for this data is administrative rather than research, it is very helpful for capturing information about populations at the time that services are provided. It also is helpful:

- In cases where program participants may not respond to a survey.
- To capture changes relevant to individuals that may be difficult to capture outside the specific service context.
- As it may provide potential comparison groups for your evaluation.
- For conducting complex statistical analyses due to large sample sizes.

Program administrative data is often collected on participant intake and exit forms, which can serve as baseline data as the program matures.

Secondary data

Unlike primary data (data collected by you, for your program), secondary data is collected by someone external to your program (e.g. national data sets, administrative or survey data collected by a different organisation such as the Australian Bureau of Statistics) which are relevant to the analysis of the outcomes of your program.

There are two forms of secondary data. One is secondary survey data or census data. When accessed in unit record form (i.e., records at the individual de-identified level), such data are very powerful tools for analysis. An example is the Household, Income and Labour Dynamics in Australia (HILDA) Survey which is a household-based panel study that has collected information about economic and social dynamics of life in Australia over the last twenty plus years.

Secondary administrative data is that held in public datasets of participants in your program or of populations similar to your program participants. It refers to public hospital records, police and court records, income support payment data and the like. The process of linking program data to public datasets ('data linkage') can provide a very rich picture of the journeys taken by program participants and the outcomes achieved by a program. The use of linked administrative data is also important in economic evaluation to assess the cost offsets (e.g., reductions in public hospital use) of a program. The process of linking data from a program to public administrative datasets can be a long and costly process and is largely done by academic research teams.

See Figure 14 on next page for some advantages and disadvantages of using these quantitative data sources.

Figure 14:

Advantages and disadvantages of quantitative data sources

SURVEYS	PROGRAM ADMINISTRATIVE DATA	SECONDARY DATA
• **Advantages:** Reach large numbers of people for a relatively low cost, and can be repeated to track changes in people's behaviours. • **Disadvantages:** Response rates, however, can be low which can jeopardise the validity of data collected.	• **Advantages:** Using this data for research allows none to capture populations who may not respond to a survey, link various information about the same individuals and identify counterfactual and control groups. • **Disadvantages:** Quality may vary due to definitional or data collection, input, management or access issues. Program data is usually collected on a participant intake form and is renewed regularly, as necessary.	• **Advantages:** Low cost, may have large sample size, are usually subject to rigorous quality control checks, and represent a good source to select counterfactuals and control post hoc for error. • **Disadvantages:** Information may be limited to data required for administrative purposes and may lack key information you need, changes to administrative procedures may change definitions and then compatibility over time (e.g. categories for age, or unemployment), quality issues for variables of less interest to administrator (e.g. address may not be updated).

QUALITATIVE METHODS

Interviews, focus groups and case studies are some commonly used methods to collect qualitative data.

Interviews

Interviews typically involve a one-on-one conversation between one person collecting data and one person talking about their experience either face-to-face, over the phone or online. Interviews allow people to talk in their own words and explore topics in-depth. They range from highly structured (standardised questions), semi-structured (a topic guides broad areas to be covered) or unstructured (narrative-style interview).

Focus groups

Focus groups are a conversation between a small group of people, facilitated by a researcher or data collector. They aim to generate discussion, debate, to provide a holistic view among the group or show a variety of opinions. Sometimes focus groups are called 'workshops', if they involve participants working on an activity together.

Case studies

Case studies are often used to illustrate good practice, provide contextual data and allow a deeper analysis of a particular outcome.[124] They can involve bringing together multiple methods of data collection and an in-depth investigation of one or a few individuals involved in the program and the people with whom they engage. The purpose is to provide particularly rich data to understand a novel situation.

See Figure 15 on next page for some advantages and disadvantages of using these qualitative data sources.

Figure 15:

Advantages and disadvantages of qualitative data sources

INTERVIEWS	FOCUS GROUPS	CASE STUDIES
• **Advantages:** Interviews can produce detailed data, helpful for establishing trust with the participant (e.g. hard to reach groups). • **Disadvantages:** Resource intensive (both time and money).	• **Advantages:** Provide the opportunity for rich data to be collected for people to "bounce" ideas off each other. Can be useful to test concepts (e.g. prior to a survey) or to explore concepts more in-depth (e.g. after a survey). • **Disadvantages:** Like interviews, focus groups can be relatively resource intensive.	• **Advantages:** Useful to understand a significant or novel situation. • **Disadvantages:** Risks include the risk of "cherry picking" - the researcher needs to ensure participants selected are likely to expose positive and negative outcomes.

Other qualitative methods

Other qualitative methods such as observation, yarning, photovoice or artwork may be useful, depending on the demographics of your key stakeholders. For example, artwork is a tool that can be used by Indigenous people to express their stories about and experiences of colonization and structural racism.[125] Artwork may be used as a form of data collection for many different components of an evaluation (e.g. problem analysis, stakeholder engagement, data collection, analysis and knowledge translation).

You should also consider using interpreters if you are collecting qualitative data from people who speak English as an additional language.

SYSTEMS- AND COMPLEXITY-BASED METHODS

There are different approaches that are used for systems thinking in outcomes measurement and evaluation.[126,127,128] Some common approaches include:

Soft systems methodology

Soft systems methodology involves analysing complex situations where there are different views about what the problem is and how it is best addressed.[129] This approach facilitates discussion between key stakeholders about the problem, with the aim of coming to consensus about changes that are desirable and achievable. Rather than focusing on the design of a framework or program, it involves a process of agreeing upon the best solution for the issue at hand. Activities such as consultation, document review, collective brainstorming and use of 'rich pictures' may be used.[130]

Critical systems heuristics

Critical systems heuristics explore systems of thought about complex issues.[131] This approach is based on a rationale that a 'true' understanding of reality is limited – and that we must engage in continual reflection in relation to our own assumptions that shape our understanding of an issue.[132] Critical systems heuristics consists of four boundary categories that each consist of a set of questions. Through working through these questions, we can obtain a deeper understanding about social problems as they allow for analysis of the mental models that lead us to understand situations in a particular way.

Social network analysis

Social network analysis (SNA) is an approach to measuring and analyzing relationships and connections between individuals and organizations.[133] The data can come from a variety of sources such as surveys, interviews, or publicly available data sets. SNA is commonly depicted in a network map in which nodes denote individuals/organisations and ties denote relationships. It can help an evaluation by analyzing how well people are working together, whether they are sharing information, and whether there is a high degree of trust between partners. Another way it can assist an evaluation is by understanding how information that is delivered in a social intervention is spreading throughout a network. It can also be used to track network outcomes that have a strong relationship component such as social support or bullying.

System dynamics

System dynamics is a methodology used for studying and managing complex feedback systems.[134] It models and analyses dynamic systems in social, managerial, economic or ecological settings. System dynamics complements systems thinking by measuring interactions and creating a picture of how a system behaves relative to time. The main goal is to help people make better decisions for strategy and policy design.

There are two basic approaches to systems dynamics. The first is approaches that map relationships and aim to understand their consequences, or develop theories about them. The second is approaches that simulate these relationships in order to understand the consequences of timing, delays and feedback associated with a program. Systems dynamics approaches are not intended to provide certainty about program effectiveness. Rather, they are used as a tool to facilitate discussion about the dynamics of the systems in which they operate.[135]

Case study: Social network analysis for evaluating partnerships

An evaluation investigated the network qualities of a partnership between a not-for-profit service provider and major utility companies in Australia.[136] The partnership aimed to address financial hardship by providing tailored services to customers experiencing difficulties paying bills and servicing loans.

The evaluation used a mixed-methods approach, including interviews, focus groups and social network analysis (SNA) to explore the development, structure and impact of the partnership. The evaluation assessed whether the partnership represented genuine collaboration or another form of joint action, revealing instrumental benefits and challenges associated with this type of partnership.

The findings revealed a high degree of trust and information sharing between partners, but a low degree of new work connections which indicated that the partnership was not shaping new ways of working and new collaborative opportunities. This SNA evaluation was useful information to help assist the development of the partnership. The partnership led to improved referral pathways and information flow between organizations, resulting in increased client referrals and improved quality of care. Based on the findings, the not-for-profit adapted some of its internal processes and capabilities to improve the partnership and project outcomes. This SNA evaluation was useful information to help assist the development of the partnership.

MIXED METHODS

Employing mixed methods for data collection (e.g. bringing different types of quantitative and qualitative data collection techniques together) **helps increase the confidence in your findings**. Mixed methods can be used concurrently (e.g. open-end interviews that explore the same themes covered in a survey) or sequentially (e.g. a focus group is used to understand key themes, which can then inform the design of a survey). Here are some questions to help you decide what type of data to collect and how:

- ***Who will you collect data about? From whom?*** This is a good time to consult (again!) your stakeholder analysis and outcomes measurement and evaluation framework. It is important to understand who the information is about and who you will ask (e.g. you may ask the individual who the program affected, but also their peers or family).
- ***What is the best instrument to collect the data, and how?*** Thinking of the characteristics of the participants/respondents and the type of information you need, assess whether a survey (face-to-face, online, mail), interview, focus group or systems-based method may be more appropriate. Also consider whether a specific interviewer/facilitator (e.g. female or male) or setting (e.g. community setting rather than institutional setting) is more appropriate.
- ***Are there any established, pre-tested instruments?*** E.g. scales for measuring certain conditions and attitudes. If there are, you must make sure you collect the data according to recommendations (face-to-face/pen and paper).
- ***Are the methods culturally appropriate?*** This may include thinking about language, norms and values. It is a good idea to consult with community representatives when developing the data collection tools.

And in the context of your program and resources:

- Consider what is a good sample size,[xii] the timing for data collection given your context (e.g. school holidays), and reimbursement for time.
- ***Staff skills to collect this data.*** Assess whether your staff is skilled to collect the respective data and whether there is need for training or outsourcing the data collection.
- Considering the ***range of data sources*** and resources (staff, skills, funding, respondents) select the ***most appropriate for your program***.

Appendix 3 presents examples of data collection plans comprising various data sources (e.g. survey, interviews, self-assessment), the type of questions the data source informs, how long the respective data collection exercise is likely to take, and things to consider when you decide to use a data collection method. Complementing the outcomes and indicators developed in Section 5, Table 6 on next page presents a snapshot of the outcomes framework for *Sport*.

xii For surveys, there are online sample calculators that based on the population size, confidence intervals and margin of error can calculate sample size that will make your survey representative of the population, for example www.surveymonkey.com or http://www.raosoft.com/samplesize.html. Sampling techniques for qualitative interviews differ and it is important to include respondents from across the population, including the whole range of characteristics (e.g. Indigenous and non-Indigenous, across all geographic areas, across age groups, genders, etc.)

Table 6:

Outcomes Framework - Outcomes, indicators, data sources and data collection time, fictional case study *Sport* program

Outcome	Indicator	Data source	Data collection time
Students understand the benefits of sport (short-term)	Proportion of students able to identify benefits of improved physical activity (quantitative)	Which of the following may be benefits of sport? (select all that apply): healthy growth of muscles and bones; better use of time; heart health; social skills; team skills; balance and coordination; better learning at school. Source: Student survey	Survey of students 3 months into the program
Students participate regularly in sport (medium-term)	Uptake of sport Absenteeism from sport	Number of students enrolled in Sport relative to total number of students in the school Absence rates – average number and proportion of absences per participant Source: Administrative data	At 6 and 12 months post program inception
Students participate regularly in physical activity (long-term)	Proportion of students engaging in at least 60 min of exercise per day	How many minutes per day are you physically active? Source: Student survey	Annual

Note: This is for illustrative purposes only and only some outcomes were included.

Case study: your job your way

The Your Job Your Way program evaluation[137] demonstrates the effectiveness of using mixed methods in data collection to comprehensively evaluate a program. The pilot program, targeting long-term unemployed youth, used a variety of data collection techniques to capture diverse perspectives and measure both vocational and non-vocational outcomes.

Key mixed-method approaches included:

- Longitudinal participant interviews: Captured youth experiences and tracked changes over time.
- Stakeholder interviews: Included insights from service providers, employers, and trusted adults.
- Administrative and output data analysis: Assessed program delivery and outcomes using quantitative measures.

A youth-friendly data collection approach was designed to minimise participant burden, ensuring the methods were inclusive and respectful of the target group. This combination of qualitative and quantitative methods provided a holistic understanding of the program's impact, including its cost-effectiveness, and informed advocacy efforts for client-centred employment interventions.

This project highlights how mixed methods can provide a richer understanding of program outcomes, especially for initiatives addressing complex social challenges.

THE DATA COLLECTION PROCESS

When undertaking data collection, ensure you consider:

- **Roles** in your program, team or organisation, their understanding and capacity to collect the data - ensure staff have the skills and time allocated to collect the data.
- **Interest and availability of participants** - your program participants are interested in supporting outcomes measurement and evaluation and available to provide you with the information you need.
- **Accuracy of data** – make sure your tools are developed to capture valid information about your evaluation questions and outcomes.
- **Relevance of data** - tools for data collection can change over time and should be revised if proven not to collect information as planned.
- **Timing and frequency** - set clear expectations about when data should be collected; this can have a significant impact on the measurement of outcomes and use of evaluation findings.
- **What is the sample size** - you should consider not only the number of participants to collect data from, but also their characteristics. Look to collect data from individuals with the same characteristics (e.g. socio-economic and demographic characteristics) as the group with which your program is engaging (i.e. the individuals surveyed are representative of their population).
- **Ensure data security and confidentiality** - data should be kept on secure servers or locations that can be accessed only by the research team in a de-identified manner, ensuring that individuals cannot be linked to their answers and their answers cannot influence their relationship with the program (refer below for more information about ethics in data collection).
- **Data sovereignty** – in outcomes measurement and evaluation affecting Indigenous people and communities, local Indigenous experts should have a say over what data is collected and how, and how data is used.

Piloting: why, when, how many?

A pilot program is a small-scale, short-term trial that helps an organisation understand how a program might work in practice. **The pilot precedes the implementation of a larger scale program and its purpose is to identify shortfalls and opportunities to improve delivery to attain the desired outcomes** for the target population. It may generate preliminary information on the extent to which intended outcomes may be achieved, although there is no direct relationship between the findings from a pilot evaluation and those from the program evaluation. Pilots are also a good opportunity to test process and learn how to better operationalise and implement the program in future.

There is no clear sample size required for a pilot study[138] as this often depends on the purpose (to validate scales, test program implementation or validity), target population, funding and time. The recommended sample size is 10-15 participants per group for feasibility studies, 25-40 participants for instrument development, or 30-40 participants per group for pilot studies comparing groups;[139] a sample that is "representative of the population and sufficiently large, respectively".[140]

It is essential to evaluate the results of pilot studies, including process and outcomes evaluation to assess the extent to which intended and unintended outcomes were achieved and whether the processes need further revision. This may be a good time to rework your planned program using your theory of change and program logic.

The pilot should:

- Be implemented according to the theory of change and program logic underpinning the program.
- Engage a sample that is representative of the population targeted by the program.
- Be evaluated to understand potential for improvement and scaling.

ETHICS AND POLITICS OF DATA COLLECTION IN OUTCOMES MEASUREMENT AND EVALUATION

Key ethical principles for outcomes measurement and evaluation (Figure 16 below) include:

- **Integrity:** professionalism, excellence (using appropriate, and proportionate methods), honesty, reliability, stewardship.
- **Beneficence:** doing no harm, protecting people from harm, managing the burden of participation, linked again to using appropriate and proportionate methods.
- **Justice:** consider the meaning of participation, not compounding inequity, being transparent about how participants are selected. Participation - or not – in outcomes measurement and evaluation activity should be independent of a person's service delivery experience. This needs to be clearly communicated to people.
- **Consent:** people need to understand what participation will mean and how their data will be used. If outcomes measurement and evaluation occurs over a long period or at various points, consent may need to be gained in an ongoing way. There are particular considerations for obtaining consent for children and young people.
- **Confidentiality:** safety of data, who can access and why, any exceptions (e.g. disclosure of threats to harm), and ensuring people are not identified in reporting products as part of knowledge translation.
- **Merit:** using sound and known methods, with quality assurance built in. Quality assurance might look like peer review, reference groups of experts or public communication.[141]
- **Value:** outcomes measurement or evaluation activities should bring value to people. 'Value' should be defined with and agreed to by the group the program aims to affect.
- **Accountability:** those undertaking the outcomes measurement and evaluation work must be accountable to their actions, and be accountable to people affected by the activity (and/or the program).

Consider the particular needs of the population you serve. You may work with people who are vulnerable, over-researched, have statutory involvement, may be fearful of saying no, or where there are cultural considerations. Consider the impact of participation for people.

Ethics needs to be considered *throughout* your outcomes measurement and evaluation journey. Whether or not you have obtained ethics approval, you should apply the above principles throughout the planning, design, data collection, data analysis and knowledge translation process.

It is important to also consider ethical requirements for evaluators:

- **Systematic enquiry:** assessment should be rigorous and include a discussion of limitations, not overclaiming.
- **Competence:** adhere to research standards and rigour, reporting should be comprehensive and accessible.
- **Integrity/honesty:** disclosure of conflict of interest; report fairly and accurately.

Figure 16:
Ethical principles for outcomes measurement and evaluation

7

ANALYSE DATA

DATA ANALYSIS

Data analysis will depend on the type of evaluation you are conducting, what data you can collect and the timing of its collection. Qualitative data is often collected at a single point in time, although 'repeat interviews' may be within the purpose of an outcomes measurement and evaluation plan. Quantitative data may be collected at a single point in time across a single or two or more groups, requiring cross-sectional analysis to allow you to identify differences between sub-groups of participants. Quantitative data collected at two or more points in time, from two or more groups requires more sophisticated statistical analysis.

We present below some key considerations for data analysis. While this will give you an overview of each type of data analysis, you should gain advice or further support for analysis in which you are not experienced.

Qualitative data analysis

Qualitative data may be audio recorded. If so, you need to transcribe the data (transfer the data to written format). There are specialized services who can do this for a cost. You will then need to code and analyse the data, and document your findings.

- **Coding data** means dividing up the data among common topics or categories that are mentioned within it, almost as if you were creating your own database. Sometimes the topics or categories are those expressed by the participants themselves in the data (explicitly or implicitly), whereas at other times the topics or categories might be pre-set and informed by the needs of the research (e.g. informed by the literature, the evaluation questions, evaluation terms of reference or outcomes framework etc).
- **Analysing data** means organising the basic topics or categories from the coding into a more sophisticated conceptual model to express the ideas contained within the whole dataset. Sometimes this process might be informed by social theory. It often means refining the names and framing of the topics and categories.
- **Coding and analysis** can be done in Word/by pen and paper/in collaborative workshops, but is more commonly done using a computer software, such as NVivo.

Braun and Clarke's (2006)[142] process for thematic coding and analysis is often used and cited as best practice – this involves:

1. Familiarising oneself with the data by reading and re-reading transcripts.
2. Generating initial codes from participants' responses.
3. Searching for themes within the initial codes.
4. Reviewing and refining the themes.
5. Defining and naming the themes.
6. Producing a write up of the findings.

It may also be appropriate (or necessary) to undertake participatory data analysis with key stakeholders. This can, firstly, support relevant stakeholders to have ownership over social change and, secondly, ensure that evaluation findings reflect what is most useful and relevant to local communities.[143]

Case study: Homelessness Deep Dive

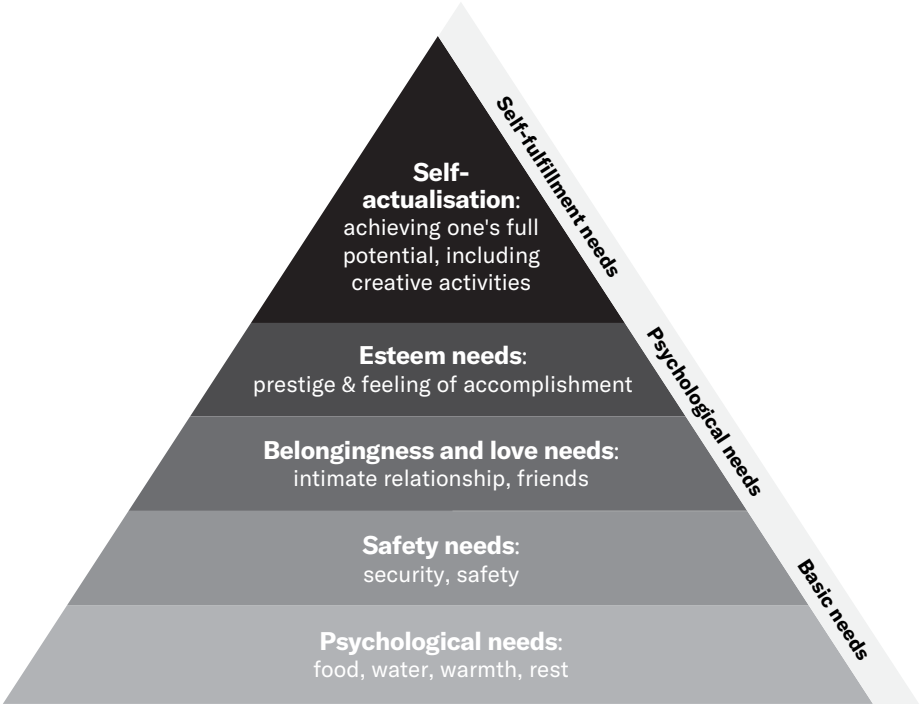
Ending Homelessness in Australia: An Evidence and Policy Deep Dive[144] was a research report that was created in partnership with the Australian Alliance to End Homelessness (AAEH) and Neami National.

Completed by researchers from CSI at the University of Western Australia (UWA) and CSI UNSW, the objectives of this research included:

1. To collate and assess the current evidence base on the state of homelessness in Australia and its key drivers.
2. To set out an evidence-informed policy and practice agenda towards ending homelessness in Australia.

One CSI UWA-based analysis included in the report involved the analysis of qualitative responses of people experiencing homelessness to the question 'what do you need to be safe and well' as included within the Australian Family Service Prioritization Decision Assistant Tool (F-SPDAT) and Australian Vulnerability Index Service Prioritization Decision Assistant Tool (VI-SPDAT). A total of 10,678 valid responses were recorded, and these were analysed according to Maslow's hierarchy of needs (see Figure 17 below).[145]

Figure 17:
Maslow's hierarchy of needs



Responses were coded manually into categories and subcategories against Maslow's hierarchy (see Figure 18 on next page).

Figure 18:
Coding against Maslow's hierarchy of needs

Category	Subcategory	Examples from data
Physiological Needs	Food/water	"Food", "Water", "Food in my belly", "Three meals a day"
	Warmth	"Warmth", "Air conditioning", "Warm place", "Warm clothes", "Clothes"
	Rest	"Sleep", "Comfortable bed", "Just to rest"
Safety Needs	Physical health	"Regular GP visits", "Bulk billed GP", "Surgery", "Pain medication"
	Mental health	"Take care of my mental health", "Mental health support", "Clear mind", "Counselling"
	Drug and alcohol	"Stay off the grog", "Stay clean", "Stay away from drugs"
	Security	"To be safe", "To be away from partner (domestic violence)", "Doors that lock", "Security for my house"

Within the research report, the data were presented in terms of the number of mentions against each subcategory under the different types of 'need'. This demonstrated which needs were particularly salient to participants (e.g. 'housing, a home and shelter' was mentioned 9,457 times).

Maslow's hierarchy of needs[146] is a theoretical framework that conceptualises people as having co-existing and holistic needs across various life domains. If needs are met, a person is better equipped to continue to meet other needs and ultimately, have their potential realised. In using a well-established framework for analysis of responses, the evaluators could substantiate their claims that people experiencing homelessness have needs that go beyond housing, and therefore responses to homelessness must consider the whole person.

Quantitative data analysis

Data collected in hard format (i.e. pen and paper) should be digitised (most often this means transferring to Excel). You can then:

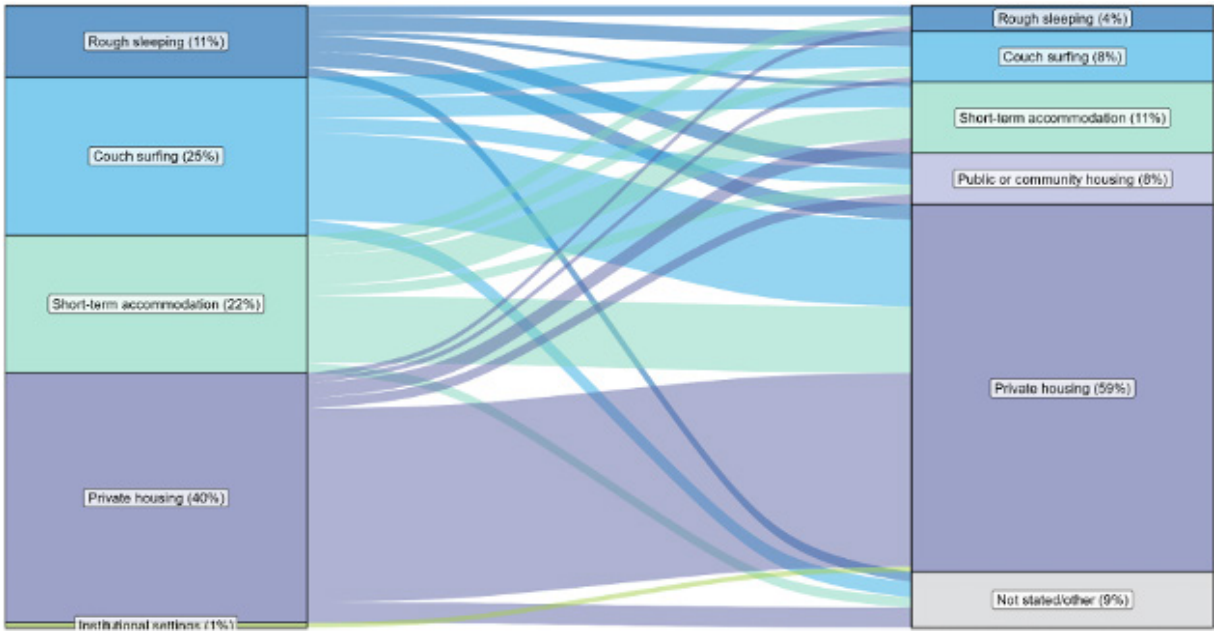
- Conduct **simple analyses**, such as descriptive statistics – these will give you a first impression about how respondents answered a question, what proportion agreed to a certain statement or how many people completed your survey.
- Conduct **complex analyses** to assess change across two or more periods of time, or differences between groups. For example:
  - Test whether the difference in one concept reported by one group (e.g. satisfaction with health, all respondents) has increased since the beginning of the program.
  - Check whether two groups are statistically different from each other (e.g. if women's satisfaction with health is significantly lower or higher than that of men) at one point in time (e.g. at the start or the end of the program).
  - Check whether the difference between two groups (e.g. men and women) has narrowed by the end of the program, compared to when the program started.

These tests, and many more, can be conducted by uploading data in statistical packages such as Stata or SPSS. Some tests can also be conducted in Excel. For example, using error bars you can conclude if observed differences (e.g. the level of satisfaction with health) have significantly changed since the start of the program or the change is due to chance. If the error bars overlap, the difference between two values is not statistically significant. See Appendix 4 for more information about using comparison groups.

Tracking a pre/post change in homelessness is not straightforward given the complexity of definitions of homelessness and the dynamic movement of people throughout various types of homelessness experiences. This quantitative analysis approach enabled a demonstration – visually – in overall positive change toward less homelessness across several different categories. This provides a summative approach to demonstrating improved outcomes.

Figure 19:

Diagram showing quantitative analysis of change in homelessness status as a result of Entrypoint



Case study: Evaluation Of Centrecare's Entrypoint Outreach Program

Centrecare's Entrypoint Outreach Program was a preventative early intervention pilot designed to divert families and individuals from homelessness into housing. It provided brief intervention (2-6 weeks) to targeted families who were newly homeless or were at immediate risk of homelessness in a tight rental market.

The program was evaluated by CSI UWA[147] according to a comprehensive outcomes matrix. A key outcome the program aimed to achieve was improvements to housing stability. The data were analysed through linking housing tenure entry and outcome data (SHS and Entrypoint's client data) and conducting an impact analysis. This is displayed in Figure 19 above, which shows a visual representation of change in housing status of Entrypoint clients as a result of the Entrypoint program.

ATTRIBUTION OR CONTRIBUTION?

Outcomes measurement and evaluation empower organisations to understand the change their activities are causing for the people they support, or the extent to which a program contributes to resolving a social problem. Distinguishing between attribution and contribution is essential. Change, especially long-term change, may be difficult to attribute to a single intervention, hence discussing ***contribution*** rather than ***attribution*** is often preferred.[148]

Contribution analysis

'Contribution analysis' is one approach that can help to understand the contribution of an intervention.[149] Steps for contribution analysis include:

1. Develop the theory of change and logic model.
2. Assess the existing evidence on your program's results (evidence that the program's activities produced the expected outputs and the expected, and unexpected, outcomes).
3. Assess the alternative explanations (the extent to which external factors may have influenced the same outcomes).
4. Assemble the narrative (why it is reasonable to assume that the actions of the program have contributed to the observed outcomes? Clarify the credibility of and weaknesses in this rationale).

Traveling through these steps involves data collection from a range of stakeholders internally (e.g. beneficiaries) and externally – people knowledgeable about the program (e.g. local community members).

Ways of determining attribution

There are some established techniques to demonstrate attribution of long-term outcomes (or impact) to a program. These are based on either measuring outcomes or impact through establishing a counterfactual such as in the RCT and QED impact evaluation designs. This can be done over a long period of time, or, if this is not possible, then it can be done through seeking the highest quality data and rigorous methodology so that one can more confidently make claims about long-term impacts based on short term outcomes data. Refer back to 'Impact evaluation' in Section 4 for outcomes measurement and evaluation designs that can help to determine attribution.

APPLYING ANALYTIC FRAMEWORKS

Once you have collected your data, applying an appropriate analytic framework is crucial for interpreting it effectively. Analytic frameworks provide structured approaches for analysing both qualitative and quantitative data, ensuring that findings are systematically derived and aligned with your evaluation goals.[150,151,152,153]

Types of analytic frameworks

Different types of analytic frameworks are suited to different types of data. Choosing the right framework depends on the type of data you have (quantitative, qualitative, or mixed methods) and the questions you want answered. Below are common types of frameworks, categorised by data type:

Quantitative analytic frameworks:
- *Descriptive statistics*: Summarise basic features of your dataset (e.g., averages or percentages).
- *Inferential statistics*: Draw conclusions about a population based on sample data (e.g., regression analysis).
- *Cost-benefit analysis*: Compare outcomes achieved against program costs to evaluate efficiency.

Qualitative analytic frameworks:
- *Thematic analysis:* Identify patterns or themes within qualitative data (e.g., interview transcripts).
- *Grounded theory:* Develop theories based on observed patterns within qualitative data.
- *Narrative analysis:* Analyse stories or personal accounts to understand individual experiences.

Mixed methods frameworks:
- *Triangulation:* Combine quantitative and qualitative data sources to cross-check findings.
- *Integration:* Synthesise insights from different types of data into a cohesive interpretation.

Steps for applying analytic frameworks

1. Clearly define your evaluation questions—what do you want to learn from the analysis?
2. Select an appropriate framework based on whether you are working with quantitative, qualitative, or mixed methods data.
3. Ensure alignment between your analytic framework and the outcomes framework established earlier in the process (see Section 5).
4. Apply the framework systematically—whether through statistical software for quantitative data or coding techniques for qualitative data.
5. Interpret results in light of your program's goals and objectives—what do these findings tell you about your evaluation questions?

ASSESSING QUALITY OF EVIDENCE

Assessing the quality of evidence is essential to ensure that the data collected during outcomes measurement and evaluation is **robust, reliable, and meaningful**. High-quality evidence allows you to make informed decisions about your program and communicate its impact confidently.[154]

Once you have applied an appropriate analytic framework to your data, it is crucial to assess whether the evidence produced meets high standards of quality.

Criteria for high-quality evidence

To assess the quality of your evidence, consider the following criteria:

- **Validity**: Does your evidence accurately measure what it is intended to measure? For example, if you are measuring changes in physical activity levels, are your indicators directly related to physical activity (e.g., step counts)?
- **Reliability**: Is your evidence consistent over time and across different groups? Reliable data should yield similar results when measured under the same conditions.
- **Relevance**: Does the evidence you've collected align with your program's goals and objectives? Ensure that the data you gather speaks directly to your evaluation questions.
- **Triangulation**: Cross-check findings by using multiple sources or methods (e.g., combining surveys with interviews) to enhance credibility and accuracy.
- **Ethical considerations**: Ensure that data collection processes respect participant rights and adhere to ethical standards, such as those outlined by NHMRC[155] and AIATSIS.[156]

Evaluating different types of evidence

Different types of evidence contribute different insights into evaluation questions and assessment of outcomes. **Quantitative data often provides measurable, generalisable findings, while qualitative data captures deep, contextual understandings.** When combined effectively, mixed methods evidence offers a comprehensive view of different aspects of programs.

Consider the following when evaluating specific types of evidence:

- **Quantitative evidence**: Check if your sample size is adequate and if statistical methods are appropriate for drawing conclusions. Consider the generalisability of your findings beyond the sample population.
- **Qualitative evidence**: Make sure that your data provides deep insights into participant experiences. Are themes well-saturated (i.e. no new information emerges from additional data)?
- **Mixed methods evidence**: If using both qualitative and quantitative data, ensure that these methods complement each other. Triangulating findings from different data sources (e.g., survey data and focus groups) can provide a more comprehensive understanding of outcomes.

Tools for assessing evidence quality

Several tools and frameworks can guide you in assessing evidence quality:

- **Checklists and rubrics:** Tools such as the "GRADE Framework"[157] or the "BetterEvaluation Rainbow Framework"[158] provide systematic approaches for evaluating the robustness of evidence.
- **IRIS+ and SROI:** Frameworks like **IRIS+**[159] (for social and environmental performance) or **Social Return on Investment (SROI)**[160] help ensure evidence meets high standards.

UNDERSTANDING BIAS

Bias can distort the findings of outcomes measurement and evaluation, leading to inaccurate conclusions about a program. **Understanding and mitigating bias is critical to ensuring that your evaluation results are credible and ethical.** This section explores common types of bias in outcomes measurement and evaluation and provides strategies for addressing them.[161,162,163]

Types of bias in outcomes measurement and evaluation

Bias can arise at various stages of outcomes measurement and evaluation, from participant selection to data collection and analysis. Understanding the different types of bias is the first step in addressing their potential influence on your findings. Below are some of the most common biases encountered in outcomes measurement and evaluation:

- **Selection bias**: Occurs when participants in your program or evaluation do not represent the broader population. For example, if only highly motivated individuals participate in the *Sport* program evaluation, their outcomes may not reflect those of less motivated students.
- **Confirmation bias**: The tendency to interpret new evidence as confirmation of pre-existing beliefs. This can occur when evaluators focus on data that supports their hypotheses while ignoring contradictory information.
- **Measurement bias**: Arises from flaws in how data is collected or measured. For example, using a poorly designed survey may lead to inaccurate responses.
- **Response bias**: Participants may provide socially desirable answers rather than truthful responses, particularly if they believe certain answers will please the evaluator.

Strategies to mitigate bias

While bias cannot be entirely eliminated, proactive strategies can significantly reduce its impact on your evaluation results. These approaches focus on designing sound evaluations, reducing bias in data collection, and interpreting findings transparently.

Addressing bias in evaluation design and data collection

Consider the following strategies:

- Use random sampling methods to ensure that participants are representative of the population you are studying. This helps reduce selection bias.
- Where possible, keep data collectors unaware of specific hypotheses or program goals to avoid influencing participant responses.
- Where appropriate, use validated tools that have been tested for reliability to minimise measurement bias.
- Engage a wide range of stakeholders—including beneficiaries—in both design and analysis phases to mitigate confirmation bias.

Addressing bias in data analysis

Bias can also affect how data is analysed. To reduce this risk:

- Be transparent about theoretical frameworks that inform your analysis, and any assumptions that are made during analysis.
- Use multiple analysts or peer review processes to cross-check interpretations.
- Report both positive and negative findings to avoid selective reporting.

8

UNDERTAKE KNOWLEDGE TRANSLATION

EFFECTIVE KNOWLEDGE TRANSLATION

In outcomes measurement and evaluation, how findings are used is as important as the findings. Potentially, **strategic use of findings can have as much positive impact as the program itself.** Effective knowledge translation supports accountability, learning and change through communicating *about* results and communicating *for* results.[164] Communicating *about* results informs stakeholders about the findings of your evaluation. Communicating *for* results (also known as 'communication for development' or 'program communication')[165] focuses on learning and improvement. This can involve various processes whereby stakeholders can actively engage with the findings, to deepen their understanding the "So What?" of the findings (see below).

Knowledge translation should also adhere to an important ethical principle – reciprocity – meaning sharing the results with participants to bring benefit to the community/group that the program is for. If you have been effective in the engagement stage, stakeholders are likely to have a strong interest in and curiosity about the findings.[166,167]

GETTING TO THE "SO WHAT?"

Knowledge translation for outcomes measurement and evaluation work should answer the question: "**So what?"**. The "So what" refers to understanding the significance, meaning and implications of the findings for the program and its context so that positive, lasting change can be achieved. While evaluators might be experts in conducting outcomes measurement and evaluation, **knowledge translation may require partnering with those who have deep contextual knowledge to create sustainable change as a result of the outcomes measurement and evaluation process**.

Some examples of getting to the "So What" include:

- Working with program staff directly on how findings might be best put to use (for example, your evaluation finds that CALD women did not access a domestic violence service; you could use your findings to run a workshop with service providers to think of ways that it can engage better with CALD communities).
- Inviting formal comment from high level staff on findings that have implications for organisations (for example, asking senior managers or executives to help with drafting recommendations in response to findings that might require action).
- Co-developing responses and actions as a result of the learnings (for example, gathering a group of teachers together to reflect on feedback from students and designing a list of easy to implement actions in response).
- Providing community-level stakeholders with evidence of success (i.e. your findings) and working with them to share their achievements and support funding applications.

The most effective communication techniques in a knowledge translation process capture attention and interest, allowing audiences to interact with the findings.[168] Tailor findings to the audience and consider:

- Accuracy, balance and fairness.
- Format of delivery.
- Level of detail.
- Level of interaction.
- Design and appearance of the publication, product or activity.
- Accessibility.

Communicating negative or sensitive findings is an important aspect of knowledge translation. **Negative findings should be seen as opportunities for internal learning to redesign a program, improve approaches to interact with clients or deliver an activity**. Results can point out groups of the target population for which a program worked as well as those for which it didn't, thus helping to identify groups or communities who may need a different approach, or a different program altogether. This can help to develop tailored activities to achieve better outcomes.

OPTIONS FOR KNOWLEDGE TRANSLATION

Strategies for knowledge translation

Successful knowledge translation aims to turn evaluation findings into actionable insights. These strategies ensure that results are communicated effectively and also integrated into decision-making processes to maximise impact, as reflected in the following strategies:[169]

- **Internal learning loops**: Establish feedback mechanisms within your organisation to ensure that evaluation results are regularly discussed at team meetings, reflected upon, and used to inform program adjustments.
- **Interactive workshops**: Host workshops with program staff and stakeholders to collaboratively interpret findings and develop action plans based on the evaluation results.
- **Co-design with stakeholders**: Engage stakeholders (including beneficiaries) in co-designing solutions based on evaluation findings. Collaborative knowledge creation, where stakeholders actively participate in generating insights from evaluation findings, ensures that the knowledge produced is practical, relevant, and more likely to be acted upon.
- **Tailor communication formats for different audiences**: To ensure that your evaluation findings are communicated effectively, it is important to tailor your communication strategies based on the needs of different audiences. Different stakeholders may require different formats or levels of detail. For example:
  - *Policymakers* may need concise **policy briefs** that highlight key takeaways.
  - *Program staff* may benefit from more detailed **reports** or **interactive workshops**.
  - *Scholarly audiences* may require **academic papers**.
  - *Wider audiences* can be reached through public **seminars, webinars, panel presentations**, **magazine articles, social media posts or community events**.[170]

Writing an evaluation report

An evaluation report is a common format for documenting outcomes measurement and evaluation findings. When writing an evaluation report, you should include at least the following sections:

- **Executive summary**: A high-level summary of the evaluation – what it did, how it was done and its key findings.
- **Introduction**: Introduce the reader to the issue that is addressed in the evaluation, its importance, as well as the program, policy or intervention that was evaluated. The description of the project program may be a separate section.
- **Outcomes measurement and/or evaluation framework**: Includes evaluation questions, scope, purpose, outcomes, methods and limitations. Including the theory of change or program logic might also be useful.
- **Evaluation findings**: Use your evaluation questions to structure how you report the findings. You will use findings from across your data sources to answer these evaluation questions and identify achievement of outcomes.
- **Conclusions and recommendations**: A high-level summary of the successes and lessons learned, as well as how findings should be used.
- **References**: The published sources you consulted throughout your evaluation.
- **Appendices**: Additional information, tables or figures that the reader can refer to for further information or clarification. It may include the evaluation plan, an outcomes measurement framework, questionnaires tools or instruments that were used for the data collection or, more detailed results (for example further disaggregated by gender, or age groups).

Tools for knowledge translation

The tools you use for knowledge translation play a crucial role in effectively communicating findings and engaging diverse audiences. **Choosing the right tools depends on the audience, the complexity of the findings, and the intended impact**. Several tools can support effective knowledge translation:[171],[172]

- **Logic models and visual program models**: Use these tools to clearly communicate how your program's activities lead to desired outcomes. These models simplify complex relationships between inputs, activities, outputs, and outcomes, making them accessible for all stakeholders.
- **Infographics**: Create infographics that visually represent key data points from your evaluation. Infographics are particularly useful for sharing findings with non-expert audiences and explaining complex findings in a simple way.
- **Dashboards**: Develop interactive dashboards that allow stakeholders to explore real-time data from your evaluation. Dashboards facilitate ongoing decision-making and continuous monitoring of key performance indicators (KPIs).
- **Case studies**: Share case studies that highlight how specific aspects of your program have led to positive outcomes. Case studies provide relatable, concrete examples that engage diverse audiences. For case studies that involve highly personal information, check you have consent to share widely. Other approaches may be more appropriate – removing or changing personal information, or creating composite case studies which involve synthesising information from several case studies to create a 'typical' story or consumer journey.

Tools such as the BetterEvaluation Rainbow Framework[173] can guide you in selecting appropriate methods for communicating findings. Similarly, frameworks like Social Return on Investment (SROI)[174] help translate social impact into monetary terms that can be easily communicated to funders and policymakers.

Embedding continuous learning

Knowledge translation should not be a one-time activity but should be part of an ongoing cycle of learning and improvement. By embedding continuous learning into your organisation's culture,[175],[176,177,178] you ensure that evaluation results are regularly revisited, lessons learned are documented, and programs are adapted based on new insights. Consider establishing formal processes for continuous learning, such as:

- Regular review meetings where teams reflect on recent evaluation findings.
- A "lessons learned" repository where staff can document insights from evaluations.
- Peer review processes where external experts provide feedback on evaluation results.

Practical steps for knowledge translation

1. Identify your target audiences and their specific information needs.
2. Choose the most appropriate formats and tools for communication.
3. Collaborate with stakeholders to ensure findings are relevant and actionable.
4. Share findings across multiple channels to reach diverse groups.
5. Follow up with stakeholders to gather feedback and evaluate the effectiveness of your knowledge translation efforts.[179]

READY, SET, GO!

Evaluation and outcomes measurement is an activity that may take from weeks to months and years to complete. It requires a good understanding of the problem that a program is looking to resolve, the context in which it is delivered and the stakeholders involved. It needs resources (people and time) and internal and/or external skills and expertise.

While it may seem difficult at times, measuring outcomes and evaluating are invaluable to understanding the impact of a program, the changes it makes to people's lives, how services can be improved.

Case study: St. Vincent De Paul – Social Housing In NSW

The Social Housing in NSW project[180] demonstrates how strategic knowledge translation can drive systemic change. CSI UNSW partnered with St. Vincent de Paul Society NSW to evaluate social housing commitments and forecast the impact of additional investment. The project used multiple strategies to ensure findings were translated into actionable insights and practical advocacy tools.

Key strategies included:

- Compelling reports: CSI UNSW produced two reports that clearly articulated the urgent need for increased investment in social housing and quantified the potential impact of building 5,000 homes annually. These findings linked proposed investments to measurable outcomes, such as reduced homelessness and housing stress.
- Advocacy support: Findings were leveraged by advocacy groups like St. Vincent de Paul to strengthen public submissions, media campaigns, and direct lobbying efforts with policymakers. The evidence-based recommendations provided a foundation for targeted advocacy efforts.
- Media amplification: Media outlets highlighted the project's data and forecasts, broadening public awareness of the housing crisis and adding pressure on policymakers to address identified gaps.
- Collaborative discussions: CSI UNSW engaged government officials and sector leaders in discussions, using the evidence to guide conversations on actionable policy solutions and systemic housing challenges.

These efforts demonstrate knowledge translation by linking rigorous data to practical recommendations and ensuring findings were accessible to a range of audiences, including policymakers, advocates, and the public. Advocacy groups reported that policymakers engaged with the project's evidence during discussions, particularly the forecast of 5,000 homes annually, which became a focal point in policy dialogues.

This project showcases how a combination of well-structured reports, targeted advocacy, media engagement, and collaborative stakeholder discussions can maximise the impact of research findings. These strategies not only amplified the visibility of critical issues but also ensured evidence informed systemic change and policy development.

9

EMBED EVALUATION AND EVALUATIVE THINKING

OUTCOMES MEASUREMENT AND EVALUATION SKILLS AND CAPABILITIES

Ensuring you have the right outcomes measurement and evaluation skills and capabilities within your organisation, or securing these externally, is essential for relevant, rigorous and reliable outcomes measurement and evaluation.

This includes considering the skills and competencies of various people across an organisation, not only those that will be undertaking outcomes measurement and evaluation. Figure 20 below and Table 7 on next page present the seven competency domains outlined within the Australian Evaluation Society (AES) Evaluator's Professional Learning Competency Framework.[181]

Figure 20:
Domains of competence from the AES Evaluator's Professional Learning Competency Framework

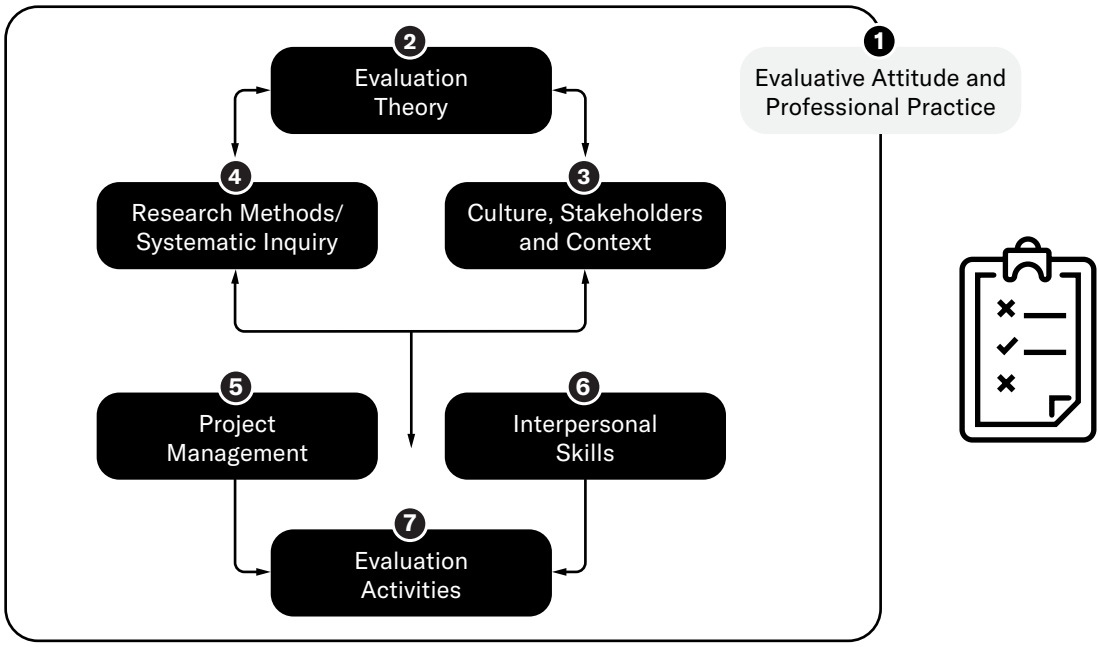


Table 7:

Description of each domain of competence from the AES Evaluator's Professional Learning Competency Framework

Competency domain	Description
1. Evaluative attitude and professional practice	Focuses on the self-reflection and on-going professional development which are critical in the broad and complex role of evaluators. This set of knowledge, skills and attitudes influence all the other competency groups.
2. Evaluation theory	Focuses on the theoretical foundations of evaluation which are distinct from other forms of inquiry.
3. Culture, stakeholders and context	Surrounded by, and works within, a multiplicity of value perspectives, including cultural, social and political. These value perspectives are embedded within the evaluand, the context within which an evaluand exists, and in the perspectives of evaluation commissioners and stakeholders. The evaluator must be cognisant of, and responsive to, such value perspectives.
4. Research methods and systematic inquiry	Knowledge and skills in research methods and systematic inquiry are essential for collecting valid and reliable data on which evaluative judgements can be based. This competency covers the knowledge and skills evaluators need to conduct systematic inquiry in an evaluation.
5. Project management	Focuses on the project management skills evaluators need to effectively negotiate, scope, manage and complete an evaluation.
6. Interpersonal skills	Focuses on the interpersonal skills evaluators need to communicate effectively with clients, consumers and other stakeholders in an evaluation.
7. Evaluation activities	The other six competency groups culminate in this set of competencies which focus on the tasks an evaluator will carry out in the course of an evaluation.

It is beneficial to continuously assess the level of skills and capabilities within your program or organisation, so that you can put existing skills and capabilities to use and address any gaps through recruitment or human resource allocation.

THE NEED FOR EVALUATIVE THINKING

Whether or not you and your team or organisation are working specifically on outcomes measurement and evaluation, and whether or not your organisation has a high level of the skills and capabilities described above, it is important to embrace and promote 'evaluative thinking' at all times. **Evaluative thinking** is:

> "a disciplined approach to inquiry and reflective practice that helps us make **sound** judgements using **good** evidence, as a matter of habit."[182]

Section 2 of the Roadmap outlined the importance of creating a culture of learning in preparation for your outcomes measurement and evaluation work. Evaluative thinking calls for individuals to adopt mindsets and reflective practices that support the use of evidence for making judgements in their work in an ongoing way.[183] It ensures that anyone involved in social impact work is using evidence in an appropriate way to inform their judgements.

While we discuss evaluative thinking at the end of the Roadmap, it is a way of thinking that should be embedded throughout an outcomes measurement and evaluation process.

APPENDICES

APPENDIX 1: USEFUL RESOURCES AND TOOLS

Below are provided a range of resources that you may wish to consult to help with your outcomes measurement and evaluation work.

Organisations with resources, tools or events for outcomes measurement and evaluation

Australian Evaluation Society (AES): https://www.aes.asn.au

Australian Centre for Evaluation (ACE): https://evaluation.treasury.gov.au

Social Impact Measurement Network Australia (SIMNA): https://simna.com.au

Better Evaluation: https://www.betterevaluation.org

Outcomes measurement readiness and decision-making

Checklist on whether you're 'outcomes measurement ready' (Compass p.39): https://assets.csi.edu.au/assets/research/the-compass-your-guide-to-social-impact-measurement.pdf

Decision-making tool for social impact measurement approaches (Compass p. 30-1): https://assets.csi.edu.au/assets/research/the-compass-your-guide-to-social-impact-measurement.pdf

Survey design

A step-by-step guide for creating more accessible surveys: https://www.alchemer.com/wp-content/uploads/2019/05/Accessibility_Ebook.pdf

How to write a survey questionnaire for evaluation: A guide for beginners: https://aifs.gov.au/resources/practice-guides/how-write-survey-questionnaire-evaluation-guide-beginners

Questionnaire design tips: https://psr.iq.harvard.edu/files/psr/files/PSRQuestionnaireTipSheet_0.pdf - https://psr.iq.harvard.edu/sites/projects.iq.harvard.edu/files/psr/files/PSRQuestionnaireTipSheet_0.pdf

Principles of survey and questionnaire design: http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Basic+Survey+Design+-+Questionnaire+Design

Interview guide design

Harvard Department of Sociology strategies for qualitative interviews: https://sociology.fas.harvard.edu/sites/g/files/omnuum1481/files/sociology/files/interview_strategies.pdf

Qualitative research: A practical guide for health and social care researchers and practitioners: Chapter 13 – Interviews - https://oercollective.caul.edu.au/qualitative-research/chapter/__unknown__-13/

Western Australian Community Impact Hub: Interviewing for data collection - https://communityimpacthub.wa.gov.au/learn-from-others/insights/interviewing-for-data-collection/

Quantitative data analysis

NCVO: Analysing quantitative data for evaluation: https://www.ncvo.org.uk/help-and-guidance/strategy-and-impact/impact-evaluation/evaluation-and-impact-reporting/how-to-analyse-quantitative-data-for-evaluation/

The Magenta Book: https://www.gov.uk/government/publications/the-magenta-book

US Centres for Disease Control and Prevention: Analysing quantitative data for evaluation: https://nhsevaluationtoolkit.net/wp/wp-content/uploads/2024/01/Analyzing-Quantitative-Data-for-Evaluation.pdf

University of Arkansas: Analyzing quantitative data - https://www.uaex.uada.edu/support-units/program-staff-development/docs/Analyzing%20Quantitative%20Data.pdf

Qualitative data analysis

Business Research Methodology – Qualitative data analysis: https://research-methodology.net/research-methods/data-analysis/qualitative-data-analysis/

Evaluation and outcomes measurement guides

Kellogg Foundation Evaluation Handbook: https://wkkf.issuelab.org/resource/the-step-by-step-guide-to-evaluation-how-to-become-savvy-evaluation-consumers-4.html

The Magenta Book. Guidance for evaluation, HM Treasury 2011: https://assets.publishing.service.gov.uk/media/5e96cab9d3bf7f412b2264b1/HMT_Magenta_Book.pdf

The Green Book. Central Government guidance on appraisal and evaluation, HM Treasury 2018: https://assets.publishing.service.gov.uk/media/6645c709bd01f5ed32793cbc/Green_Book_2022__updated_links_.pdf

Office of Social Impact Investment (2018). Technical guide: outcomes measurement for social impact investment proposals to NSW Government: https://www.nsw.gov.au/sites/default/files/2023-05/technical-guide-outcome-measurement-2018-july.pdf

Western Australian Community Impact Hub: https://communityimpacthub.wa.gov.au/

NSW Health – Planning and Managing Program Evaluations: https://www.health.nsw.gov.au/research/Publications/program-evaluations.pdf

CDC Program Evaluation Framework: https://www.cdc.gov/evaluation/php/evaluation-framework/index.html

University of Minnesota CYFAR. Types of evaluation: https://cyfar.org/node/131

Australian Institute of Family Studies: Planning an evaluation: https://aifs.gov.au/resources/practice-guides/planning-evaluation

CSIRO (2020). Impact evaluation guide: https://www.csiro.au/-/media/About/Files/Our-impact-framework/CSIROImpactEvaluationGuide_WEB.pdf

APPENDIX 2: TECHNICAL APPENDIX

Table 8:
Indicator selection criteria

Technical	
Specific	The level of clarity and detail in what the indicator is trying to measure, its key terms and variables.
Validated	The evidence to support that the indicator measures what it intends to measure. For example, whether the indicator has been tested in a controlled study or validated through consensus amongst practitioners and/or experts.
Reliable	The degree to which an indicator produces consistent results over time.
Comparable	The degree to which the indicator is comparable across spatial areas, groups and against existing benchmarks or target levels.

Contextual	
Important	How important and useful do you think this data is?
Acceptable	How comfortable do you think participants would be to be asked this information?
Appropriate	How relevant do you think this indicator is for the participants' situations?
Feasibility	How practical would it be to collect this information from participants?

Table 9:

Data collection approaches matrix

Approach	The types of questions the method answers	Burden on respondent	Things to consider	Generalizability
Sample survey	"Approximately how prevalent is the outcome/issue in your target population"? "What factors are associated with the outcome/issue and how strong are those associations"?	Low – 10-20 mins	• What level of quality does your data need to have? How accurate does your data need to be? • Who are your desired respondents? • How will you contact people and ask them to respond? • How will you encourage people to respond? • What will the question items be? • How will you make sure the question items are not biased or confusing to people? • How will you get the questionnaire to people, and get their responses back? • How will the data be entered into a database? • Where will the database be stored? • When the data come in, how will you make sure they have been entered accurately? • How will the data be analysed? • What data analysis skills and expertise do you need?	Depends on the accuracy of your data. So, if collected with high amount of rigor, the data are generalizable to target population, with some margin of error
Census survey	"How prevalent is the outcome/issue in your target population"? "What factors are associated with the outcome/issue and how strong are those associations"?	Low – 10-20 mins		Depends on the accuracy of your data. So, if collected with high amount of rigor, the data are generalizable to entire target population
Self-assessment	"How much do individuals think they have changed, re outcome or issue"?	Low – 10-20 mins		Depends on the accuracy of your data and whether it took a sample or census approach
Standardised test assessment	"How much have individuals changed, in their knowledge and skills"?	High – Time + Test preparation/ anxiety		Depends on the accuracy of your data and whether it took a sample or census approach
Interviews	"Approximately how prevalent is the issue"? "How appropriate is the program"? "How has the program impacted your life"? "What might be causing this issue or facilitating these outcomes (because you have no preconceived ideas)"? "How could the program be improved"?	Medium – 30-90 mins	• Who will conduct interviews? • What level of quality does your data need to have? How accurate does your data need to be? • Who are your desired respondents? • How will you protect participants safety, privacy and confidentiality?	Depends on the accuracy of your data and whether it took a sample or census approach

Approach	The types of questions the method answers	Burden on respondent	Things to consider	Generalizability
Focus groups	"How appropriate is the program"? "How has the program impacted your life"? "What might be causing this issue or facilitating these outcomes (because you have no preconceived ideas)"? "What terms does the target population use to describe the issue/outcome and its causes"? "How could the program be improved"?	Medium – 30-90 mins	• Who will lead the focus group? • How will you contact people and ask them to participate? • How will you encourage people to participate? • What questions will you ask them? • How will you make sure the question items are not biased or confusing to people? • How will you ensure that the person leading the interview or focus group will encourage honest, unbiased responses? • How will you ensure participants safety is prioritised? • How will you protect participants safety, privacy and confidentiality? • How will the verbal responses be recorded? • How will themes from the responses be identified? • How will other aspects of the data be analysed? • What data analysis skills and expertise do you need?	Depends on the accuracy of your data. Even if collected with high amount of rigor, the data have limited generalizability to target population
Observation	"What is happening in the immediate spatial and social context surrounding people as they work toward an outcome or handle an issue"?	None to low	• What level of quality does your data need to have? How accurate does your data need to be? • What is the target of observation? • How will entry to the location of observation be gained? • How will you ensure that the person observing is doing so in an unbiased manner? • How will observations be recorded? • How will observations be analysed?	Depends on the accuracy of your data. Even if collected with high amount of rigor, the data are generalizable only to a portion of the target population, and only in that setting
Other qualitative methods, such as observation, artwork or photovoice	"What is your experience of the issue/program"? "How has the program impacted your life"?	Depends on method	• Who will pay for resources etc required (e.g. camera, paint, paper)? • How much time will be allowed? • How will the data be integrated into the analysis?	Depends on recruitment and sampling but unlikely to be generalizable to the broader population (as it focuses on individual experiences). Cross-cutting themes may be developed

Approach	The types of questions the method answers	Burden on respondent	Things to consider	Generalizability
Administrative data (e.g., case management data)	"How many people are in the system"? "Who accesses the program"? "What are their characteristics"?	Depends on administrative recording process, but generally low to medium	• How were data collected from respondents? • What questions were asked? • How are the data stored and organized?	Depends on the accuracy of the data. Even if collected with high amount of rigor, the data are generalizable only to the population in the administrative system
Document review (content analysis)	"What is the individual or organization publicly saying about the outcome or issue"?	None	• Who created the document(s) and who does that person or those people represent? • How will the document(s) be stored? • How will themes or data be identified?	Not generalizable to broader population
Secondary data	Depends on the method of data collection used by the original collector	None	• How were data collected from respondents? • What questions were asked? • How are the data stored and organized? • How were respondents identified, selected and recruited? • How was the questionnaire designed, worded and delivered? • Were the data entered accurately? • How are the data stored and organized? • What is the data analysis approach needed?	Depends on method originally used and the accuracy of the data

APPENDIX 3: DATA COLLECTION PLAN TEMPLATES

A data collection plan systematically guides you through data required for the key components of your outcomes measurement and evaluation framework. There are numerous ways to structure a data collection plan, but the general idea is to take your evaluation questions and/or outcomes and turn it into a plan you can implement to ensure the data you collect aligns with what you set out to understand.

Figure 21 and Figure 22 below (sourced from Kaleveld et al., 2020)[184] show examples of how data collection plans can be structured.

Figure 21:

Example of a data collection plan (structured according to evaluation question)

Evaluation questions – typically no more than 5 questions	Focus of evaluation	How will data be collected and when?	How will data be stored and who will have access?	Who is responsible?
E.g. How effective is 'Together 4 Tea' at reducing social isolation in 65–80 year olds?	Identifying any reduction of prevalence and/or severity of social isolation amongst the target group.	Focus groups and surveys will be conducted to gather baseline data prior to first meetings. Data will be collected in the same manner after the first round of four meetings, and again after nine months.	Data will be stored on the local government's secure drive and will be accessible only to the evaluators.	The community development manager will be responsible for oversight of the evaluation plan, however the community development project officer will be responsible for designing and implementing the evaluation.

Figure 22:

Example of a data collection plan (structured according to evaluation question and outcome)

Evaluation questions (Developed from program logic)	Focus of evaluation (What do we want to know?)	Indicators and measures (How will we know it?)	Data Collection (How will data be collected?)	Frequency (When will data be collected?)	Responsibility (Who will collect the data?)
E.g. Have 'Together 4 Tea' participants experienced a reduction in the severity of social isolation?	What are participants' experiences of social isolation like now, in comparison to prior to the program?	Self-reporting decreased experience of social isolation.	Baseline data collected via postal survey.	Baseline data will be collected prior to the initial meeting.	Community development project officer. (Note: initial meeting facilitators are not to conduct focus groups due to risk of confirmation bias.)
		Improved health indicators over time.	Follow up data collected by postal surveys and focus groups.	Follow up data will be collected after the first round and again nine months after the first round.	

APPENDIX 4: ESTABLISHING COMPARISON GROUPS

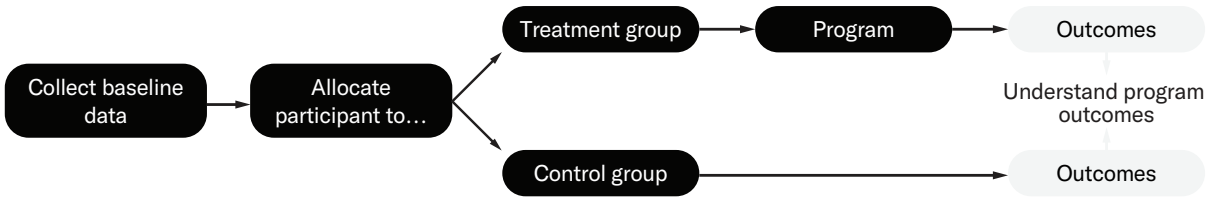
A core characteristic of high-quality evidence is the use of a 'comparison group' to help demonstrate the effectiveness of an initiative. This would involve participation in your program of two groups – one group that has been exposed to the program, compared to a second group that has not been exposed to the program. Through the presentation of information in this way, one can draw conclusions about the extent to which an intervention makes a difference, in a demonstrable way.

There are several ways to achieve a comparison group, through various designs, as described below.

Randomised Controlled Trials (RCTs)

RCTs are a scientific experiment looking to minimise forms of bias when assessing change due to a program.[185] The population eligible to participate in a program is randomly divided into two groups. One group participates in the program (treatment group) while the other does not (control group). Outcome data is collected from both groups prior to the program (at baseline) and again during, at the end, and/or some time after the program has been completed. Analysing the differences between the two groups prior and post program allows comparisons to be made, which can inform conclusions about the outcomes that can be attributed to the program (see Figure 23 below).

Figure 23:
Diagram showing comparison of outcomes between treatment and control group in an RCT



RCTs assume that individuals across the two groups are identical in all aspects except for their program participation, presenting several issues to consider for good design[xiii] including:

- It can disturb what would naturally occur.
- Individuals selected in the treatment group may act differently than they otherwise would, due to selection. Those in the control group may also act differently as 'resentful demoralisation' (The Hawthorn Effect).
- Some in the control group may seek out or receive the treatment in one form or another (substitution bias, contamination or crossover effects).
- RCTs are generally small and run as demonstrations so may fail to capture community wide effects (scale bias).
- RCTs focus on effectiveness outcomes alone but no explanation of why the program works (causality).
- Whether the findings of the experiment can be reproduced in 'real world' context (external validity).
- Ethical issues: those in the control group may be considered disadvantaged (while unproven, intervention is hypothesised to bring advantages to participants); resources are usually limited, and some high-need potential clients may miss out on potentially life-saving intervention (RCTs use random allocation, as opposed to a needs-based allocation).[xiv]
- Ensuring the groups are identical - e.g. using non-volunteers as comparison group for a program where participation is voluntary may prove problematic, because refusal of participation may signal different characteristics in participants.
- Maintaining contact with the comparison group - while participants in the treatment group are in contact with staff, this is not the case for the comparison group and unless circumstances allow for repeated data collection (pre and post program), data collection from the comparison group might compromise the analysis.
- High cost.

Quasi Experimental Designs (QEDs)

QED is a similar experiment to RCT with the exception of the randomisation of the sample e.g. in QEDs inclusion in the treatment group may be subject to eligibility (such as income threshold) and hence the validity of the study may be jeopardised. The availability of a control group may also be problematic; a control group needs to be equivalent to the treatment group on a range of relevant characteristics. The researcher compares the outcomes of the two groups prior to the program then again post-program.

When designing a QED ensure that you are aware of potential threats to the integrity of the data[186]:

- **Selection-history threat:** events between pre- and post-test that the groups experience differently (e.g. growing up, students in the control group went to a certain school, while those in treatment did not).
- **Selection-maturation threat:** different rates of normal growth between pre-test and post-test for the groups (hence achieving the outcome may be the natural course, for example, students engaging in certain, age-appropriate activities).
- **Selection-testing threat:** a *differential* effect between groups on the post-test of taking the pre-test (they learn differently from pre-test).
- **Selection-instrumentation threat:** *differential* change in the test used for each group from pre-test and post-test.
- **Selection-mortality threat:** *differential* non-random dropout between pre-test and post-test.
- **Selection-regression threat:** different rates of regression to the mean in the two groups (e.g. if initially one of the groups has members with extremely low scores).

The analysis of change between groups can be done in statistical packages such as SPSS or Stata, as well as Excel. Figure 24 on next page illustrates the change in two groups and presents how error bars can be used in Excel to assess whether the change is statistically significant.

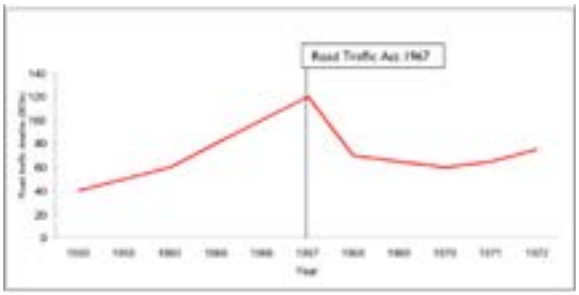
---

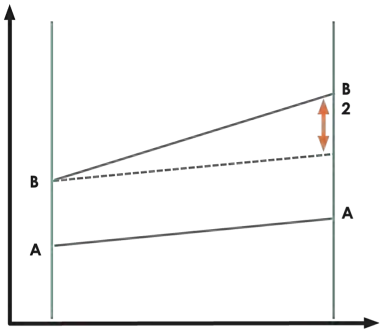
xiii See Magenta Book175, Paper 7, p.7.7 for further discussion of RCT and limitations.

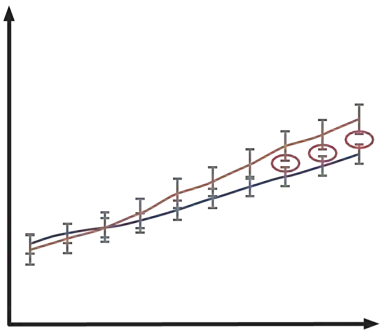
xiv There are several techniques to overcome this issue, for example phased introduction (a program is rolled out in waves), intermittent applications (when interventions are very short-term and may be repeated at different intervals across different groups), or accidental delays (such as implementing nation-wide policies).

Figure 24:

Examples: Assessing change using comparison groups







Interrupted time series (a form of QED), whereby observations of a constant variable are examined over time in order to look for 'interruptions' to the series of observations that may have resulted from a program (or policy). In the example showing a clear reduction in road traffic deaths at the time of the introduction of the Road Traffic Act[aci], evaluators would need to rule out other possible causes such as taxes on road usage or petrol, or changes to the price of alcoholic beverages (Source: Green (1997)[187] from Baker (2000)[188] referenced in HM Treasury (2007), p.7:26).[189]

Difference-in-difference (Group B participated in the program (treatment group), while group A didn't (control group). Steps: plot values at time 1 and time 2 for treatment and control groups. Visualise (the dotted line) what would have happened, had group B not taken part in the program. The difference noticed, B2 to C, can be attributed to receiving the treatment.

Error bars can be used as a first exploration of change, for the same group or between two or more groups. This is a function in Excel. If the error bars overlap, then the difference between the two values is not statistically significant (and rather due to chance). If the bars do not overlap, then you can conclude that significant change has occurred.

APPENDIX 5: USE OF ARTIFICIAL INTELLIGENCE IN OUTCOMES MEASUREMENT AND EVALUATION

While there are differing definitions for what constitutes Artificial Intelligence (AI), one definition includes ***the ability for a non-biological entity to achieve complex goals***.[190] In other words, AI allows computer programs and systems that can perform tasks usually done by humans.[191] In the past decade or so there has been exponential growth of AI, with its application expanding into nearly all areas of life. It has coincided with exponential growth of big data due to rapid evolution of new technologies and reduced costs of data storage.

Artificial intelligence is created through a process called 'machine learning'. ML starts when a large amount of 'training' data is entered into a computer. The computer is then asked to identify patterns that structure this data to generate an algorithm. Then, when users give the computer data, it applies this algorithm to the data it was provided with. The computer then applies the algorithms it has already developed to the new data so that it creates new outputs.[192]

Generative AI refers to the ability of machines to create new information without human intervention. They are able to generate content without the need for explicit programming. One example of Generative AI is ChatGPT, which is a web-based question and answer chat interface that users can have conversations with.

In theory, AI can be used to conduct any task relating to outcomes measurement and evaluation. As with all methodologies and tools, AI can be used for evaluation on the condition that you are transparent in any reporting about what activities you used AI for, how you used it, and its limitations. For outcomes measurement and evaluation work AI can be used for the following stages of an outcomes measurement and evaluation process shown in Table 10 on next page according to the Rainbow Framework[ak] (adapted from Australian Evaluation Society, 2024).[193]

Table 10:

Potential uses for AI in outcomes measurement and evaluation

Stage of outcomes measurement and evaluation	Potential uses of AI
Establish purpose	• Understanding the possible causes of a social issue. • Understanding how a systems-thinking lens can be applied to a social issue. • Identifying potential levers for change.
Plan	• Identifying and analysing stakeholders. • Estimating a budget. • Identifying and mitigating potential ethical issues.
Clarify program design	• Developing a theory of change or program logic. • Identifying possible unintended outcomes.
Understand what to evaluate	• Developing key evaluation questions. • Identifying the appropriate type of evaluation.
Develop an outcomes measurement and evaluation framework	• Identifying outcomes and indicators. • Identifying relevant datasets.
Collect data	• Describing sampling and recruitment. • Designing data collection methods and tools e.g. survey design, participatory activities. • Collecting data. • Transcribing data.
Analyse data	• Coding data. • Visualising results. • Applying rubrics to different datasets. • Explaining evaluative reasoning.
Undertake knowledge translation	• Reframing reports to particular audiences (e.g., creating summary reports). • Structuring arguments in a different way (e.g. bullet points or narrative).

This outline of potential uses however, is quite idealised, and as with any technology there are both opportunities and risks associated with its use.[194,195,196]

Opportunities associated with the use of AI for outcomes measurement and evaluation include:

- Democratisation of knowledge and power for those who traditionally cannot access it: for instance, through prompting stakeholders who might not be researchers to examine in more depth any areas of the findings that are relevant for them, and to produce tailored reports for their stakeholder groups.
- Greater process efficiencies: such as less time spent on tasks that can easily be completed by AI resulting in more time available for knowledge translation and other strategic work.
- New jobs and skills: leading to new opportunities for workforce training and development.

Risks include:

- Privacy and confidentiality breaches: AI presents an ethical 'grey zone' when sharing personal stories or anecdotes as it is unclear where this raw data is stored and how it might be used in the future.
- Disinformation caused by 'hallucinations' (errors): AI is only as good as what it is taught and where there are gaps in understanding, it can produce unreliable or erroneous information.
- Training data are curated by people and therefore are subject to existing patterns of discrimination: findings can reflect entrenched social prejudices and therefore amplify discrimination towards marginalised peoples (e.g. recruitment bias, bias in ways data is aggregated, bias in interpretation of findings and attributions for 'success' or 'failure' or how they are defined).
- Intellectual property: issues relating to plagiarism and ownership throw doubt over how ethical its use is, and hinders our ability to comprehensively and accurately reference and attribute theories and concepts.
- AI falls short in one of the core principles of research: whereby one transparently outlines methodology to create findings that are replicable. Conversely, AI is more of a 'black box' where one cannot properly interrogate how findings were developed (i.e., the processes by which how raw data was turned into a finding or idea). This reduces rigour and confidence in findings.
- Lack of consideration of context: AI is built on mathematical formulas, and reflects what is statistically the 'most likely' sequences of words and ideas. It cannot effectively take into account or consider nuances of local contexts, or specific 'time and place' influences.
- We are still understanding what skills and competencies are needed for responsible and effective use of AI – i.e., what makes for a good prompting, what level of coding literacy is needed for optimal use.
- There are also more general concerns about job losses/structural change and disruption caused by AI, which has led some companies to pause their use on it entirely until some of these issues are resolved.

In summary, AI presents the world of evaluation with many new opportunities to assist with the more mundane tasks of evaluation work, creating more time and space for exploring, interrogating and reflecting strategically on the findings. Evaluators must exercise caution in how they use AI models, for ethical reasons such as avoiding inputting any raw and identifiable data in language model systems. Also, evaluators must conduct critical reviews of any 'findings' generated by AI, considering the following:

- Could findings reflect entrenched biases?
- Do they adequately consider context?
- Can we still describe the processes through which conclusions were reached, so our presentation of findings is rigorous?
- Do the processes and outputs that AI were engaged in represent the values/valuing that are core to this program and its evaluation?

So far, there has been equal measure of curiosity and caution within the community of evaluators about the potential of AI. One guideline that evaluators keep coming back to is in the use of AI for any evaluation processes, it is important to be aware of the pitfalls, and to 'have a human in the loop' at all stages.

The Australian Evaluation Society often runs discussion groups, seminars and conference presentations on this topic, as it is a rapidly evolving area. One way to stay abreast of developments and frameworks for using AI in evaluation is by following the AES' engagement in this emerging field.

GLOSSARY

- **Activities:** The processes or actions that produce the desired outputs and ultimately outcomes. In essence, activities describe 'what we do'.
- **Artificial intelligence:** The theory and development of computer programs and systems that can perform tasks usually done by humans.
- **Attribution:** Attribution is the extent to changes can be considered a direct result of a program or intervention.
- **Baseline:** The initial information collected about the condition or performance of subjects prior to the implementation of a program or intervention, against which progress can be compared at strategic points during and at completion of the program.
- **Benchmarking:** A process of measuring an outcome, change or performance against 'reference points' from another established program or national measures. For example, measuring education outcomes from a local program against state-level education outcomes.
- **Blue marble evaluation:** An holistic approach to evaluation that emphasises the interconnectedness of human and ecological systems. Blue marble evaluation puts forth the need for evaluators to consider the broader environmental, social, and economic impacts of a program in the context of sustainable development and the health of the planet.
- **Continuous quality improvement:** Progressive incremental improvement of operations, outcomes, systems and processes.
- **Cultural responsiveness:** The practice of paying particular attention to social and cultural factors in interactions with people from different cultural and social backgrounds. People working in culturally responsive settings continuously self-reflect to examine and unlearn their own assumptions about race and culture.
- **Cultural safety:** Refers to Indigenous people feeling like their cultures are valued, their experiences are validated and their knowledges and skills are recognised. Cultural safety also refers to absence of all forms of racism in Indigenous peoples experiences of policy development, evaluation, research and service delivery.
- **Data collection plan:** A structured tool that helps to define program objectives, activities, outputs, indicators and data sources. Helps to facilitate systematic evaluation of programs.
- **Data linkage:** Identifying and combining information from corresponding records (e.g. individuals or cases) from different data sources (e.g. program administrative data and public administrative datasets).
- **Developmental evaluation:** An approach that involves asking evaluative questions and gathering information to provide feedback and support decision-making along an emergent path in the context of innovation. Primarily focusing on adaptive learning rather than accountability, its purpose is to provide real-time feedback and generate learning.

- **Economic evaluation:** The assessment of the efficiency of a program by comparing outcomes achieved against the costs of the program. Techniques include cost-benefit analysis and cost-effectiveness analysis.
- **Evaluand:** The object of an evaluation, usually a program, policy, initiative or service etc.
- **Evaluation:** A systematic process of assessing and understanding the merit, worth or significance of a program, by combining evidence and values. In an evaluation, social research procedures are systematically applied to assess the conceptualisation, design, implementation, and utility of programs or initiatives.
- **Evaluation framework (or plan):** A document explaining all the elements of how you plan to evaluate a program according to your selected approach. An evaluation framework focuses on what needs to be evaluated from a strategic perspective, whereas an evaluation plan includes operational steps for how you will evaluate something. For example, your plan or framework may include evaluation questions, design, methodology, governance, risk and an outcomes measurement framework (if relevant).
- **External evaluation:** An evaluation conducted by an individual or organisation that sits external to the organisation that delivers the program being evaluated.
- **Formative evaluation:** Evaluation that occurs during a program's implementation, with the aim of improving design and performance.
- **Goals:** Longer-term aspirations your organisation has for the future. Goals indicate where your organisation's efforts are directed.
- **Impact:** The longer-term sustained social, economic, and/or environmental effects or consequences of a program.
- **Impact evaluation:** The assessment of the extent to which long- term, sustained changes resulted from the program activities. This type of evaluation is more likely to influence policy.
- **Indicators:** Indicators are measurable markers that show whether progress is being made on a certain condition or circumstance. Different indicators will be needed to determine how much progress has been made toward a particular goal, output, or outcome.
- **Inputs:** Resources inserted into a program for its establishment and implementation. Examples include money, staff, time, facilities, equipment.
- **Internal evaluation:** An evaluation conducted by an individual or organisation that sits internally within the organisation that delivers the program being evaluated.
- **Knowledge translation:** A dynamic and iterative process of analysing, sharing, exchanging and applying knowledge to improve programs and services.
- **Lived experience:** Refers to the firsthand knowledge that individuals gain through direct involvement with social issues or programs.
- **Methods:** The tools, techniques and processes used to measure outcomes and evaluate.
- **Mission:** Describes 'the business' of an organisation or a program. It describes how the future described through its 'vision' will be achieved.
- **Objectives:** More tangible, specific and measurable aspirations that aim to achieve an organisation's or program's goal.
- **Outcome:** An outcome can be both the results/effects expected by implementing a program/initiative/ strategy and the changes that occur in attitudes, values, behaviours or conditions. Changes can be immediate, intermediate or long-term.

- **Outcomes evaluation:** The assessment of the changes resulting from the implementation of a program, policy or other initiative It includes both intended and unintended outcomes for a range of stakeholders engaging in a program or initiative.
- **Outcomes framework:** A collection of selected outcomes, indicators to measure those outcomes and the data sources necessary to quantify or qualify those indicators. It also includes the timing for data collection to measure change in each outcome.
- **Outcomes measurement:** A systematic way to assess the extent to which a program has achieved its intended results.
- **Outcome statement:** A specific, measurable statement that defines what a program is aiming to achieve.
- **Outputs:** The direct products or services resulting from a program or initiative's activities. For example, the number of people, places, supports or activities your program has produced.
- **Qualitative data:** Data that seeks to understand how the world is understood, interpreted and experienced by individuals, groups and organisations (usually through the eyes of people being studied and in natural settings). It unpacks the 'why', is often flexible, relative and subjective. Qualitative data is usually text or narrative.
- **Quantitative data:** Data that seeks to explain something by using numerical data: how many, much, often; change etc. They are highly structured and based on theory/evidence and usually objective, but can also capture subjective responses (e.g. attitudes, feelings etc).
- **Philanthropy:** The planned and structured giving of time, information, goods and services, voice and influence, as well as money, to improve the wellbeing of humanity and the community.
- **Policy:** A deliberate course or principle that governs actions and decisions in different contexts.
- **Positionality:** Refers to personal values and identities that influence how one engages with and understands the world as a result of the social and political context in which one lives.
- **Problem analysis:** A process that helps understand the entrenched nature of social issues, identify the 'root causes' and helps map potential solutions.
- **Process evaluation:** The investigation of the extent to which a program or initiative was implemented as planned. It helps understand why changes occurred.
- **Program:** A coordinated number of related projects and other activities, that aim to deliver established goals or objectives.
- **Program logic:** A visual representation of how a program will achieve its goals, including the short-, medium- and long-term outcomes. It comprises a detailed representation of inputs, activities, outputs, outcomes and impact.
- **Realist evaluation:** An approach to evaluation that seeks to understand how and why certain programs achieve different outcomes for different people using a context, mechanism and outcomes model.
- **Shared measurement:** Shared measurement is a process for coordinating consistent measurement and evaluation (e.g. how to measure, what indicators to use). It is also a tool for collecting and measuring results consistently across groups.

- **Social finance:** Investment into enterprises or initiatives that aim to secure social, environmental and (in some cases) financial returns. Social finance is implemented by organisations such as government and philanthropists who wish to contribute to social change.
- **Social impact:** The intended and unintended social consequences, positive and negative, of programs (services, policies, plans, projects) and any social change processes invoked by these.
- **Social impact assessment:** The processes of analysing, monitoring and managing social impact.
- **Social network analysis:** An evaluation method that measures and analyzes relationships and connections between individuals and organizations using data from surveys, interviews or network mapping.
- **Social purpose ecosystem:** All those who deliver, or support the delivery of, services or programs to improve the lives of individuals or communities. It is an increasingly mixed ecosystem where government and for-profit, not-for-profit, and philanthropic organisations, as well as individuals, work either separately or together towards improved social outcomes.
- **Statistical significance:** The likelihood that the relationship between two or more variables or two or more groups is not due to chance. For example, it helps to understand if the difference in an outcome (improved mental health) between girls and boys is due to chance or it was caused by an intervention.
- **Stakeholders:** Any group or individual who can affect, or is affected by, an organisation or its activities. Also, any individual or group that can help define value propositions for the organisation.
- **Summative evaluation:** Evaluation that occurs at the end of a program's implementation or when it is well-established enough to make assessments about criteria such as appropriateness, effectiveness, efficiency and sustainability. Summative evaluations are often intended to inform decisions about ongoing investments and/or scaling.
- **Systems thinking:** Understanding a whole system – e.g. the social system – by examining the links and interactions between the components.
- **Theory based evaluation:** An approach to evaluation that is based on a theory of change or program logic that explains mechanisms of change. The evaluation process aims to test whether the theory works. The role of the evaluator is to produce evidence that shows what has changed, and through which mechanisms.
- **Theory of change:** An explicit theory or model of how a program will achieve the intended or observed outcomes. It articulates the hypothesised causal relationships between a program's activities and its intended outcomes and identifies how and why changes are expected to occur. In doing so, the theory of change comprises a change model (the changes the program intends to achieve) and an action model (the activities that will lead to those changes). A theory of change must be plausible, doable and testable.
- **Vision:** An organisation's statement of its overall ideal and the ultimate goal of its operations. May be framed as a description of what the future should look like. Provide strategic direction and facilitates decision-making.

REFERENCES

1 Ramia, I., Powell, A., Stratton, K., Stokes, C., Meltzer, A., & Muir, K. (2021). *Roadmap to outcomes measurement. Your step-by-step guide to planning, measuring and communicating social impact.* Centre for Social Impact.

2 Muir, K., & Bennett, S. (2014). *The Compass: Your guide to social impact measurement.* The Centre for Social Impact.

3 Better Evaluation. (2025). *What is evaluation?* https://www.betterevaluation.org/getting-started/what-evaluation

4 HM Treasury. (2011). The Magenta Book: Guidance for evaluation. London: HM Treasury. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220542/magenta_book_combined.pdf

5 Vanclay, F. (2003). International Principles For Social Impact Assessment. *Impact Assessment and Project Appraisal*, 21, 5–12.

6 Smart, J. (2020). *Program planning and evaluation guide.* Australian Institute of Family Studies. https://aifs.gov.au/sites/default/files/publication-documents/2006_planning_and_evaluation_step_by_step_0.pdf

7 Rittel, H.W., & Webber, M.W. (1973). Dilemmas in a general theory of planning. Policy Sciences, 4, 155–169. doi:10.1007/BF01405730.

8 Hu, D., Zhou, S., Crowley-McHattan, Z.J., Liu, Z. (2021). Factors that influence participation in physical activity in school-aged children and adolescents: A systematic review from the social ecological model perspective. *International Journal of Environmental Research and Public Health*, 18, 3147. https://doi.org/10.3390/ijerph18063147

9 Clearinghouse for Sport. (2024). *Factors influencing sport participation*. https://www.clearinghouseforsport.gov.au/kb/factors-influencing-sport-participation

10 Meadows, D.H. (2008). *Thinking in Systems.* Earthscan UK.

11 Bronfenbrenner, U., & Morris, P. A. (2006). The Bioecological Model of Human Development. In R. M. Lerner & W. Damon (Eds.), *Handbook of child psychology: Theoretical models of human development* (pp. 793–828). John Wiley & Sons.

12 Waters Center for Systems Thinking. (2020). Habits of a systems thinker (2020 Edn.). https://ttsfilestore.blob.core.windows.net/ttsfiles/habits-single-page-2020.pdf#page=1&view=FitH

13 Williams, B. & Imam, I. (2007). *Systems Concepts in Evaluation: An Expert Anthology.* EdgePress.

14 Williams, B., & Hummelbrunner, R. (2011). *Systems concepts in action: A practitioners toolkit.* Stanford University Press.

15 Gates, E.F., Walton, M., Vidueira, P. & McNall, M. (2021). Introducing systems- and complexity- informed evaluation. *New Directions for Evaluation*, 2021(170), 13–25. https://doi.org/10.1002/ev.20466

16 Dass, S., Muir, K., & Rangan, K. (2020). Nonprofit Boards: It Is Time to Lift Your Gaze and See the System. Working Paper 21-058. Harvard Business School. https://ssrn.com/abstract=3737045

17 Elias, A.A., & Cavana, R.Y. (2000). Stakeholder Analysis for Systems Thinking and Modelling. In. Victoria University of Wellington.

18 Centre for Homelessness Impact, RSM, & Cordis Bright. (2025). *Systems-wide evaluation of homelessness and rough sleeping: Preliminary findings.* UK Ministry of Housing, Communities & Local Government. https://www.gov.uk/government/publications/systems-wide-evaluation-of-homelessness-and-rough-sleeping-preliminary-findings

19 Wheelen, T.L., & Hunger, J.D. (2008). *Strategic Management and Business Policy* (15th ed.) Pearson.

20 Hull, C.E., & Lio, B.H. (2006). Innovation in non-profit and for-profit organizations: Visionary, strategic, and financial considerations. *Journal of Change Management*, 6(1), 53–65.

21 David, F.R. (2009). *Strategic management: concepts and cases* (12th ed.). Pearson Education.

22 Marjolin, A., Muir, K., & Powell, A. (2015). *The Travel Companion: Your guide to working with others for social outcomes.* The Centre for Social Impact.

23 Organisation for Economic Co-operation and Development – Development Assistance Committee. (2010). *Glossary of Key Terms in Evaluation and Results Based Management.* OECD-DAC. http://www.oecd.org/development/peer-reviews/2754804.pdf

24 Mayne, J. (2009). Building an evaluative culture: The key to effective evaluation and results management. *The Canadian Journal of Program Evaluation*, 24(2), 1–30.

25 Conley-Tyler, M. (2005). A fundamental choice: Internal or external evaluation? *Evaluation Journal of Australasia*, 4(1-2), 3–11.

26 Wholey, J.S. (2015). Exploratory Evaluation. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation* (pp.81-99). Jossey-Bass.

27 Paton, R., Emerson, J., & Nicholls, A. (2015). *Social finance.* Oxford University Press.

28 Bollen, P., & Sainty, E. (2022). *A guide to outcomes contracting and social impact bonds*. Social Ventures Australia. https://www.socialventures.com.au/sva-quarterly/a-guide-to-outcomes-contracting-and-social-impact-bonds/

29 Schwandt, T.A., & Gates, E.F. (2021). *Evaluating and valuing in social research.* The Guilford Press.

30 Mertens, D.M. (2009). *Transformative Research and Evaluation.* Guildford Press.

31 Milner, H.R. (2007). Race, Culture, and Researcher Positionality: Working Through Dangers Seen, Unseen, and Unforeseen. *Educational Researcher*, 36(7), 388-400.

32 Coram, V., Lester, L., Tually, S., Kyron, M., McKinley, K., Flatau, P., & Goodwin-Smith, I. (2022). *Evaluation of the Aspire Social Impact Bond: Final report.* Centre for Social Impact, Flinders University, Adelaide and Centre for Social Impact, University of Western Australia, Perth, https://doi.org/10.25916/202z-ey67

33 Mayne, J. (2009). Building an evaluative culture: The key to effective evaluation and results management. *The Canadian Journal of Program Evaluation*, 24(2), 1–30.

34 Stewart, J. (2014). *Developing a culture of evaluation and research.* Australian Institute of Family Studies. https://aifs.gov.au/sites/default/files/publication-documents/cfca28_0_0.pdf

35 Mayne, J. (2009). Building an evaluative culture: The key to effective evaluation and results management. *The Canadian Journal of Program Evaluation*, 24(2), 1–30.

36 New Philanthropy Capital. (2022). *The Code of Good Impact Practice.* https://www.thinknpc.org/resource-hub/the-code-of-good-impact-practice/

37 Khalifa, M. A., Gooden, M. A., & Davis, J. E. (2016). Culturally responsive school leadership: A synthesis of the literature. *Review of Educational Research*, 86(4), 1272–A311. https://doi.org/10.3102/0034654316630383

38 McMillan F. (2013). Culturally Responsive Health Care [Position Paper]. Indigenous Allied Health Australia. https://iaha.com.au/wp-content/uploads/2015/07/IAHA_POLICY_CRHC2015-web.pdf

39 Powell, A., Muir, K., Marjolin, A., Etuk, L., Reeve, R., Ramia, I., Voola, A., Raymond, C., & Fatimah, A. (2019). *Financial Inclusion Action Plan Evaluation.* Centre for Social Impact University of New South Wales. https://www.csi.edu.au/research/financial-inclusion-action-plan-evaluation/

40 Knowledge Institute on Child and Youth Mental Health and Addictions. (2024). *Clearer insights, greater impact: A program evaluation toolkit.* Knowledge Institute on Child and Youth Mental Health and Addictions. https://www.cymha.ca/Modules/ResourceHub/?id=f5e9e97c-46da-4e7c-937c-6e924e1d7cfd

41 Callis, Z., Flatau, P., & Seivwright, A. (2017). *Progress in Outcomes Measurement: Findings from the Second Annual Outcomes Measurement in the Western Australian Community Sector Survey.* Bankwest Foundation.

42 Krick, T., Forstater, M., Monaghan, P., & Sillanpää, M. (2005). *From Words to Action: The Stakeholder Engagement Manual.* AccountAbility, United Nations Environment Programme, Stakeholder Research Associates Canada Inc. https://stakeholderresearch.com/wp-content/uploads/sra-2005-words-to-action-stakeholder-engagement-02.pdf

43 United Nations. (2007). *Declaration on the Rights of Indigenous Peoples* (A/RES/61/295). United Nations General Assembly. https://documents.un.org/doc/undoc/gen/n06/512/07/pdf/n0651207.pdf

44 Muir, K., & Bennett, S. (2014). *The Compass: Your guide to social impact measurement.* The Centre for Social Impact.

45 Fitzpatrick, J.L., Sanders, J.R., & Worthen, B.R. (2011). *Program Evaluation: Alternative approaches and practical guidelines* (4th ed.). Pearson.

46 Wallerstein N., Duran B., Oetzel J. & Minkler M. (Eds.). (2018). Community-based participatory research for health: Advancing social and health equity (3rd ed.). Jossey-Bass.

47 Tew, J. (2006). Understanding Power and Powerlessness: Towards a Framework for Emancipatory Practice in Social Work. Journal of Social Work, 6(1), pp.33-51.

48 Fraser, H., & MacDougall, C. (2017). Participatory Action Research with Indigenous Peoples.

49 Wallerstein N., Duran B., Oetzel J. & Minkler M. (Eds.). (2018). Community-based participatory research for health: Advancing social and health equity (3rd ed.). Jossey-Bass.

50 Smith, L. (2021). Decolonizing methodologies: Research and Indigenous peoples (3rd edn). Zed Books.

51 Beresford, P. (2013). *Beyond the usual suspects: Towards inclusive user involvement.* Shaping our lives. https://lx.iriss.org.uk/sites/default/files/resources/beresford.pdf

52 Chambers, R. (1997). *Whose Reality Counts? Putting the First Last.* Practical Action Publishing.

53 Fetterman, D. M., Kaftarian, S. J., & Wandersman, A. (2015). *Empowerment evaluation: Knowledge and tools for self-assessment, evaluation capacity building, and accountability* (2nd Ed.). SAGE Publications. https://doi.org/10.4135/9781483387079

54 Chouinard, J. A., & Cousins, J. B. (2015). The Journey from Rhetoric to Reality: Participatory Evaluation in a Development Context. *Educational Assessment, Evaluation and Accountability*, 27, pp.5-39.

55 Cousins, J.B., & Whitmore, E. (1998). Framing Participatory Evaluation. *New Directions for Evaluation*, 80, pp.5-23.

56 Fetterman, D. M., Kaftarian, S. J., & Wandersman, A. (2015). *Empowerment evaluation: Knowledge and tools for self-assessment, evaluation capacity building, and accountability* (2nd Ed.). SAGE Publications. https://doi.org/10.4135/9781483387079

57 Smith, L. (2021). *Decolonizing methodologies: Research and Indigenous peoples* (3rd edn). Zed Books.

58 Denzin, N., Lincoln, Y., & Smith, L. (2008). *Handbook of Critical and Indigenous Methodologies.* SAGE Publications, Inc. https://doi.org/10.4135/9781483385686

59 Batterham, D., Tually, S., Coram, V., McKinley, K., Kolar, V., McNelis, S., & Goodwin-Smith, I. (2023). *Crisis accommodation in Australia: now and for the future* (AHURI Final Report No. 407). Australian Housing and Urban Research Institute Limited. https://www.ahuri.edu.au/research/final-reports/407

60 Gollan, S., & Stacey, K. (2021). *Australian Evaluation Society First Nations Cultural Safety Framework*. Australian Evaluation Society. https://www.aes.asn.au/images/AES_FirstNations_Cultural_Framework_finalWEB_final.pdf

61 Kukutai, T., & Taylor, J. (2016). *Indigenous Data Sovereignty: Toward an agenda*. ANU Press and the Centre for Aboriginal Economic Policy Research.

62 Ibid.

63 National Health and Medical Research Council (Australia), Australian Research Council, and Universities Australia. (2023). National Statement on Ethical Conduct in Human Research. National Health and Medical Research Council. https://www.nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research-2023

64 AIATSIS. (2020). *Code of Ethics for Aboriginal and Torres Strait Islander Research.* Canberra: AIATSIS.

65 Australian Evaluation Society (2013). *Guidelines for the Ethical Conduct of Evaluations.* Australian Evaluation Society. https://www.aes.asn.au/images/stories/files/membership/AES_Guidelines_web_ v2.pdf

66 Patton, M.Q. (2012). Budgeting for Developmental Evaluation (DE). https://www.betterevaluation.org/sites/default/files/Budgeting_for_Developmental_Evaluation.pdf

67 da Costa, P. (2012). *Study on Communicating Evaluation Results.* Prepared for the OECD Informal Network of DAC Development Communicators.

68 Conley-Tyler, M. (2005). A fundamental choice: Internal or external evaluation? *Evaluation Journal of Australasia*, 4(1-2), 3–11.

69 Rogers, P.J. (2008). Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions. *Evaluation*, 14(1), 29–48.

70 Goldsworthy, K. (2021). *What is theory of change?* Australian Institute of Family Studies. https://aifs.gov.au/resources/practice-guides/what-theory-change

71 Connell, J.P., & Kubish, A.C. (1998). *Applying a theory of change approach to the evaluation of comprehensive community initiatives: Progress, prospects, and problems.* The Aspen Institute.

72 Harries, E., Noble, J., & Hodgson, L. (2014). *Creating your Theory of Change: NPC's practical guide.* New Philanthropy Capital (NPC). https://www.thinknpc.org/resource-hub/creating-your-theory-of-change-npcs-practical-guide/

73 Coryn, C.L.S., Noakes, L.A., Westine, C.D., & Schröter, D.C. (2011). A Systematic Review of Theory-Driven Evaluation Practice From 1990 to 2009. *American Journal of Evaluation*, 32(2), 199–226.

74 Norton, M.I., Mochon, D., & Ariely, D. (2012). The IKEA effect: When labor leads to love. *Journal of Consumer Psychology*. 22(3), 453–460.

75 Department of Human Services. (2003). *Measuring health promotion impacts: A guide to impact evaluation in integrated health promotion.* Victorian Government. https://www.betterevaluation.org/sites/default/files/measuring_hp_impacts.pdf

76 Australian Agency for International Development. (2000). *The Logical Framework Approach.* Commonwealth of Australia. https://archive.org/details/AusAIDLogicalFrameworkGuide/AusAID%20Logical%20Framework%20Guide/mode/2up?view=theater

77 Bennett, S., Reeve, R., Muir, K., Marjolin, A., & Powell, A. (2016). *Orienting your journey: An approach for indicator assessment and selection.* Centre for Social Impact, University of New South Wales.

78 W.K. Kellogg Foundation. (2017). The step-by-step guide to evaluation: How to become savvy evaluation consumers. W.K. Kellogg Foundation. https://wkkf.issuelab.org/resource/the-step-by-step-guide-to-evaluation-how-to-become-savvy-evaluation-consumers-4.html

79 Innovation Network. (n.d.). *Logic model workbook.* Washington. https://innonet.org/media/logic_model_workbook_0.pdf

80 Office of Social Impact Investment. (2018). *Technical Guide: Outcomes Measurement for social impact investment proposals to NSW Government*. NSW, Australia: OSII. https://www.osii.nsw.gov.au/assets/office-of-social-impact-investment/Technical-Guide-Outcome-measurement-2018- July.pdf

81 Rogers, P.J. (2008). Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions. *Evaluation*, 14(1), 29–48.

82 Better Evaluation. (2025). *What is evaluation?* https://www.betterevaluation.org/getting-started/what-evaluation

83 OECD. (2021). *Applying Evaluation Criteria Thoughtfully.* OECD Publishing. https://doi.org/10.1787/543e84ed-en

84 Scriven, M. (1991). Evaluation thesaurus (4th Ed.). Sage Publications.

85 Rogers, P., & Woolcock, M. (2023). *Process and implementation evaluations: A primer.* CID Faculty Working Paper No. 433 [May 2023]. Centre for International Development at Harvard University. https://bpb-us-e1.wpmucdn.com/websites.harvard.edu/dist/c/104/files/2023/05/2023-05-cid-wp-433-process-and-implementation-evaluation.pdf

86 Hutchinson, C., Lester, L., Coram, V., Flatau, P., & Goodwin-Smith, I. (2024). *Beyond the bottom line: Assessing the social return on investment of a disability-inclusive social enterprise.* Centre for Social Impact at Flinders University and Centre for Social Impact at University of Western Australia.

87 Zappalà, G. (2020). Meaningful evaluation: A holistic and systemic approach to understanding and assessing outcomes. *Cosmopolitan Civil Societies: An Interdisciplinary Journal*, 12(2–3), 1–19. https://doi.org/10.5130/ccs.v12.i2-3.7034

88 Kania, J., & Kramer, M. (2013). Embracing emergence: How collective impact addresses complexity. *Stanford Social Innovation Review*, January, 1-7. https://ssir.org/articles/entry/social_progress_through_collective_impact#

89 Kania, J., Kramer, M., & Russell, P. (2014). Strategic Philanthropy for a Complex World. *Stanford Social Innovation Review*, Summer, 25-37. https://ssir.org/articles/entry/strategic_philanthropy

90 Zappalà, G. 2020. Meaningful evaluation: A holistic and systemic approach to understanding and assessing outcomes. *Cosmopolitan Civil Societies: An Interdisciplinary Journal*, 12(2–3), 1–19. https://doi.org/10.5130/ccs.v12.i2-3.7034

91 Bogomolova, S., Eyers-White, D., Tan, P.J., Jardine, B., Mackenzie, C., & Goodwin-Smith, I. (2022). *"... it gave me the oomph to go and do it ... or I'd still be sitting in the house 24/7." Community Connections program evaluation report.* Centre for Social Impact, Flinders University. https://researchnow-admin.flinders.edu.au/ws/portalfiles/portal/70318033/CCP_EvaluationReport_CHANGES_v6.pdf

92 Patton, M. Q. (2011). *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use.* Guilford Press.

93 Patton, M. Q. (2008). *Utilization-focused evaluation* (4th Ed.). Sage Publications.

94 Marchal, B., van Belle, S., van Olmen, J., Hoerée, T., & Kegels, G. (2012). Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems research. *Evaluation*, 18(2), 192–212. https://doi.org/10.1177/1356389012442444

95 Dozois, E., Langlois, M., & Blanchet-Cohen, N. (2010). *A practitioner's guide to developmental evaluation.* The J.W. McConnell Family Foundation and the International Institute for Child Rights and Development. https://www.betterevaluation.org/sites/default/files/DE%2520201%2520EN.pdf

96 Westley, F., Zimmerman, B., & Patton, M. Q. (2006). *Getting to maybe. How the world is changed.* Random House.

97 Pawson, R. & Tilley, N. (1997). *Realistic Evaluation.* Sage.

98 Wong, G., Westhorp, G., Manzano, A., Greenhalgh, J., Jagosh, J., & Greenhalgh, T. (2016). RAMESES II reporting standards for realist evaluations. *BMC Medicine*, 14(1), 96. https://doi.org/10.1186/s12916-016-0643-1

99 Marchal, B., van Belle, S., van Olmen, J., Hoerée, T., & Kegels, G. (2012). Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems research. *Evaluation*, 18(2), 192–212. https://doi.org/10.1177/1356389012442444

100 Caló, F., Roy, M. J., Donaldson, C., Teasdale, S., & Baglioni, S. (2019). Exploring the contribution of social enterprise to health and social care: A realist evaluation. *Social Science & Medicine*, 222, 154–161. https://doi.org/10.1016/j.socscimed.2019.01.007

101 American Evaluation Association. (2018). Principles for effective use of systems thinking in evaluation: Systems in evaluation TIG. https://www.systemsinevaluation.com/wp-content/uploads/2018/10/SETIG-Principles-FINAL-DRAFT-2018-9-9.pdf

102 Patton, M.Q. (2016). A transcultural global systems perspective: In search of blue marble evaluators. *Canadian Journal of Program Evaluation*, 30(3), pp. 374-390.

103 Zappalà, G. 2020. Meaningful evaluation: A holistic and systemic approach to understanding and assessing outcomes. *Cosmopolitan Civil Societies: An Interdisciplinary Journal*, 12(2–3), 1–19. https://doi.org/10.5130/ccs.v12.i2-3.7034

104 Joyce, A., Elmes, A., Campbell, P., Moussa, B., Suchowerska, R., Barraket, J., & Carey, G. (2022). The health and well-being impacts of a work integration social enterprise from a systems perspective. *Health Promotion International*, 37(1), daab052. https://doi.org/10.1093/heapro/daab052

105 Joyce, A., Dabrowski, A., Aston, R., & Carey, G. (2017). Evaluating for impact: What type of data can assist a health promoting school approach? *Health Promotion International*, 32(2), 403–410. https://doi.org/10.1093/heapro/daw034

106 Rowe, A. K. (2009). Potential of integrated continuous surveys and quality management to support monitoring, evaluation, and the scale-up of health interventions in developing countries. *The American Journal of Tropical Medicine and Hygiene*, 80(6), 971–979.

107 Ibid.

108 Joyce, A., Dabrowski, A., Aston, R., & Carey, G. (2017). Evaluating for impact: What type of data can assist a health promoting school approach? *Health Promotion International*, 32(2), 403–410. https://doi.org/10.1093/heapro/daw034

109 Meadows, D. (1999). *Leverage points: Places to intervene in a system.* Sustainability Institute.

110 Mrazek, P., Biglan, A., & Hawkins, J. (2005). *Community-Monitoring Systems: Tracking and Improving the Well-being of America's Children and Adolescents.* National Science Foundation and The Society for Prevention Research.

111 Clavier, C., & de Leeuw, E. (2013). *Health Promotion and the Policy Process.* Oxford University Press.

112 Sabatier, P. (2019). *Theories of the Policy Process* (2nd Ed.). Routledge. https://doi.org/10.4324/9780367274689

113 Joyce, A., Green, C., Kearney, S., Leung, L., & Ollis, D. (2019). Alignment and political will: Upscaling an Australian respectful relationships program. *Health Promotion International*, 34(5), 892–901. https://doi.org/10.1093/heapro/day034

114 Kessler, R.C., Andrews, G., Colpe, L.J., Hiripi, E., Mroczek, D.K., Normand, S.L.T., Walters, E.E., & Zaslavsky, A.M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine*, 32(6), 959–976.

115 Gutuskey, L. (2022). *Centering Equity in Program Evaluation* (OPRE Report 2022-211). Administration for Children and Families. https://www.acf.hhs.gov/sites/default/files/documents/opre/centering_equity_program_evaluation_feb2023.pdf

116 Doran, G.T. (1981). There's a S.M.A.R.T. Way to Write Management's Goals and Objectives. *Management Review*. 70, 35–36.

117 Bennett, S., Reeve, R., Muir, K., Marjolin, A., & Powell, A. (2016). *Orienting your journey: An approach for indicator assessment and selection.* Centre for Social Impact, University of New South Wales.

118 Wilson, E., Campain, R., & Brown, C. (2021). The Community Services Outcomes Tree. An Introduction. Centre for Social Impact, Swinburne University of Technology.

119 Barnes, E., Hartley, C., & Parker, R. (2024). *Together Home Program Evaluation.* Centre for Social Impact, University of New South Wales. https://assets.csi.edu.au/assets/Together-Home-Evaluation-Final-Report-October-2024.pdf

120 Australian Government. (2023). Measuring What Matters: Australia's First Wellbeing Framework. Commonwealth of Australia. https://treasury.gov.au/sites/default/files/2023-07/measuring-what-matters-statement020230721_0.pdf

121 ARACY – Australian Research Alliance for Children and Youth. (2024). What's in the nest? ARACY. https://www.aracy.org.au/wp-content/uploads/2024/10/Whats-In-The-Nest-2024-Update.pdf

122 United Nations. (2024). The 17 Goals. United Nations. https://sdgs.un.org/goals

123 ARACY – Australian Research Alliance for Children and Youth. (2024). *What's in the nest?* ARACY. https://www.aracy.org.au/wp-content/uploads/2024/10/Whats-In-The-Nest-2024-Update.pdf

124 Denzin, N.K., & Lincoln, Y.S. (2013). *The Landscape of Qualitative Research* (4th ed.). SAGE Publications.

125 Emmerton, R. & Giselsson, K. (2024). Indigenous art as decolonising truth-telling: Battle Mountain Memorial. *Journal of Sociology*, 60(4), 741–759. https://doi.org/10.1177/14407833241255153

126 Capra, F., & Luisi, P. (2014). *The Systems View of Life: A Unifying Vision.* Cambridge: Cambridge University Press.

127 Williams, B. & Hummelbrunner, R. (2011). *Systems Concepts in Action: A Practitioner's Toolkit.* Stanford University Press.

128 Reynolds, M., & Holwell, S. (2010). *Systems Approaches to Managing Change: A Practical Guide.* Springer.

129 Checkland, P. (1984). Systems thinking in management: the development of soft systems methodology and its implications for social science. In H. Ulrich & G. J. B. Probst (Eds.), *Self-Organisation and Management of Social Systems* (pp. 94-104). Springer.

130 Barbrook-Johnson, P., & Penn, A. (2022). *Systems mapping.* Palgrave MacMillan.

131 Hutcheson, M., Morton, A. & Blair, S. (2024). Critical systems heuristics: A systematic review. Systemic Practice and Action Research, 37, pp. 499-514.

132 Ulrich, W. (2003). Beyond methodology choice: critical systems thinking as critically systemic discourse. *Journal of the Operational Research Society*, 54(4), 325–342. https://doi.org/10.1057/palgrave.jors.2601518

133 Scott, J. (2017). *Social network analysis* (4th Ed.). SAGE.

134 System Dynamics Society. (2024). *What is system dynamics?* https://systemdynamics.org/what-is-system-dynamics/

135 Williams, B., & Harris, B. (2005). *Systems dynamics methodology.* W.K. Kellogg Foundation. https://www.bobwilliams.co.nz/ewExternalFiles/AESSD.pdf

136 Moran, M., Joyce, A., Barraket, J., MacKenzie, C., & Foenander, E. (2016). What Does 'Collaboration' Without Government Look Like? The Network Qualities of an Emerging Partnership. *Australian Journal of Public Administration*, 75(3), 331–344. https://doi.org/10.1111/1467-8500.12195

137 Ramia, I, Meltzer, A., Moffatt, J., Powell, A., & Barnes, E. (2020). *Your job your way - Final evaluation report.* Centre for Social Impact and yourtown. https://assets.csi.edu.au/assets/research/Your-Job-Your-Way-Evaluation-Report.pdf

138 Johanson, G.A., & Brooks, G.P. (2010). Initial Scale Development: Sample Size for Pilot Studies. *Educational and Psychological Measurement*. 70(3), 394–400.

139 Hertzog, M.A. (2008). Considerations in determining sample size for pilot studies. *Research in Nursing and Health*, 31(2), 180–191.

140 Johanson, G.A., & Brooks, G.P. (2010). Initial Scale Development: Sample Size for Pilot Studies. *Educational and Psychological Measurement*. 70(3), 394–400.

141 Mayne, J. (2009). Building an evaluative culture: The key to effective evaluation and results management. *The Canadian Journal of Program Evaluation*, 24(2), 1–30.

142 Braun, V., & Clarke, C. (2006). Understanding thematic analysis in psychology. *Qualitive Research in Psychology*, 3(2), 77–101.

143 Kogen, L. (2024). Participatory data analysis in social change work: empowering, practical, or both? *Qualitative Research*, 0(0). https://doi.org/10.1177/14687941241288207

144 Flatau, P., Lester, L., Seivwright, A., Teal, R., Dobrovic, J., Vallesi, S., Hartley, C., & Callis, Z. (2021). *Ending Homelessness in Australia: An evidence and policy deep dive.* Centre for Social Impact, University of New South Wales, University of Western Australia and Swinburne University. https://assets.csi.edu.au/assets/research/Ending-Homelessness-in-Australia-An-Evidence-and-Policy-Deep-Dive.pdf

145 Maslow, A.H. (1943). A Theory of Human Motivation. *Psychological Review*, 50, 370-396.

146 Ibid.

147 Flatau, P., O'Shaughnessy, D., Lester, L., Kaleveld, L., & Lester, R. (2023). Supporting Families Facing Homelessness: Centrecare's Entrypoint Outreach Program. Centre for Social Impact University of Western Australia, LotteryWest and Centrecare. https://assets.csi.edu.au/assets/CSI-UWA-Centrecare-Report-Final.pdf

148 Mayne, J. (2001). Addressing Attribution Through Contribution Analysis: Using Performance Measures Sensibly. *Canadian Journal of Program Evaluation*. 16(1), 1–24.

149 Ibid.

150 Alomoto, W., Niñerola, A., & Pié, L. (2022). Social Impact Assessment: A Systematic Review of Literature. *Social Indicators Research*, 161, 225–250. https://doi.org/10.1007/s11205-021-02809-1

151 Mertens, D.M. (2020). Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods. SAGE.

152 Creswell, J.W., & Plano Clark, V.L. (2017). *Designing and Conducting Mixed Methods Research.* SAGE.

153 Patton, M.Q. (2015). *Qualitative Research & Evaluation Methods: Integrating Theory and Practice.* SAGE.

154 Masset, E., White, H., & Waddington, H. (2018). *CEDIL Guidelines for Good Impact Evaluation Practice.* https://cedilprogramme.org/publications/goodimpact/

155 National Health and Medical Research Council (Australia), Australian Research Council, and Universities Australia. (2023). National Statement on Ethical Conduct in Human Research. National Health and Medical Research Council. https://www.nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research-2023

156 AIATSIS. (2020). Code of Ethics for Aboriginal and Torres Strait Islander Research. AIATSIS.

157 Atkins, D., Best, D., Briss, P. A., Eccles, M., Falck-Ytter, Y., Flottorp, S., Guyatt, G. H., Harbour, R. T., Haugh, M. C., Henry, D., Hill, S., Jaeschke, R., Leng, G., Liberati, A., Magrini, N., Mason, J., Middleton, P., Mrukowicz, J., O'Connell, D., Oxman, A. D., Phillips, B., Schünemann, H. J., Edejer, T., Varonen, H., Vist, G. E., Williams, J. W., Jr., & Zaza, S. (2004). Grading quality of evidence and strength of recommendations. BMJ, 328(7454), 1490. https://doi.org/10.1136/bmj.328.7454.1490

158 Better Evaluation. (2022). Rainbow Framework. Better Evaluation. https://www.betterevaluation.org/frameworks-guides/rainbow-framework

159 The Global Impact Investing Network. (2024). IRIS+. The Global Impact Investing Network. https://iris.thegiin.org/

160 Social Value International (2020). Social Return on Investment (SROI) Guide. Social Value International. https://socialvalueuk.org/resources/sroi-value-map/

161 Eriksen, S. (2018). Impact Evaluations, Bias, and Bias Reduction: Non-experimental methods, and their identification strategies. University of Groningen. https://pure.rug.nl/ws/portalfiles/portal/67762546/Chapter_1.pdf

162 Rossi, P.H., Lipsey, M.W., & Freeman, H.E. (2019). Evaluation: A Systematic Approach (8th Ed.). SAGE.

163 Better Evaluation. (2020). Bias Reduction. Better Evaluation. https://www.betterevaluation.org/methods-approaches/methods/bias-reduction

164 da Costa, P. (2012). Study on Communicating Evaluation Results. Prepared for the OECD Informal Network of DAC Development Communicators.

165 da Costa, P. (2012). Study on Communicating Evaluation Results. Prepared for the OECD Informal Network of DAC Development Communicators.

166 Canadian Institutes of Health Research. (2012). Guide to Knowledge Translation Planning at CIHR: Integrated and End-of-Grant Approaches. https://cihr-irsc.gc.ca/e/45321.html

167 Graham, I. D., Logan, J., Harrison, M. B., Straus, S. E., Tetroe, J., Caswell, W., & Robinson, N. (2006). Lost in translation: Time for a map? The Journal of Continuing Education in the Health Professions, 26(1), 13-24.

168 Torres, R.T., Preskill, H.S., & Piontek, M.E. (2005). Evaluation strategies for communicating and reporting: enhancing learning in organizations (2nd ed.). Sage Publications.

169 Nutley, S., Walter, I., & Davies, H.T.O. (2007). Using Evidence: How Research Can Inform Public Services. Policy Press.

170 Ibid.

171 Davies, H.T.O., & Nutley, S. M. (2008). Learning more about how research-based knowledge gets used: Guidance in the development of new empirical research. William T. Grant Foundation. https://wtgrantfoundation.org/wp-content/uploads/2015/10/Guidance-in-the-Development-of-New-Empirical-Research.pdf

172 Better Evaluation. (2022). Rainbow Framework. Better Evaluation. https://www.betterevaluation.org/frameworks-guides/rainbow-framework

173 Ibid.

174 Social Value International. (2020). Social Return on Investment (SROI) Guide. https://socialvalueuk.org/resources/sroi-value-map/

175 Education Development Center. (2019). Building a Culture of Continuous Improvement: Guidebook and Toolkit. Education Development Center. https://www.edc.org/sites/default/files/uploads/EDC-Building-Culture-Continuous-Improvement.pdf

176 Tan, W., & Tan, J. (2023). How to Develop Continuous Learners. MIT Sloan Management Review. https://sloanreview.mit.edu/article/how-to-develop-continuous-learners/

177 Sessa, V. I., & London, M. (2015). Continuous Learning in Organizations: Individual, Group, and Organizational Perspectives. Psychology Press.

178 Patton, M. Q. (2011). Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use. Guilford Press.

179 Levin, B. (2008). Thinking about knowledge mobilization. International Journal of Learning, 16(5), 1-12.

180 Barnes, E., Writer, T., Hartley, C., & Wearring, A. (2022). Social Housing in New South Wales. Centre for Social Impact University of New South Wales. https://www.csi.edu.au/research/social-housing-in-new-south-wales/

181 Australian Evaluation Society. (2013). Evaluators professional learning competency framework. Australian Evaluation Society. https://www.aes.asn.au/evaluation-resources/evaluator-competencies

182 Department of Education. (2023). Evaluative thinking. New South Wales Government. https://education.nsw.gov.au/teaching-and-learning/professional-learning/pl-resources/evaluation-resource-hub/evaluative-thinking

183 Buckley, J., Archibald, T., Monica, H., & Trochim, W. (2015). Defining and teaching evaluative thinking: Insights from research on critical thinking. American Journal of Evaluation, 36(3), 375-388.

184 Kaleveld, L., Atkins, N., Flatau, P., & Mollinger-Sahba, A. (2020). Measuring our impact: Evaluation framework for measuring the impact of community development work across local government in Western Australia. Centre for Social Impact University of Western Australia and Local Government Professionals Australia.

185 HM Treasury. (2007). The Magenta Book: guidance notes for policy evaluation and analysis. Background Paper 7: Why do social experiments? Government Social Research Unit.

186 Trochim, W.M.K. (n.d.). Multiple Group Threats. http://www.socialresearchmethods.net/kb/intmult.php

187 Greene, W.H. (1997). Econometric Analysis. Prentice Hall Press.

188 Baker, J.L. (2000). Evaluating the impact of development projects on poverty: A handbook for practitioners. The World Bank.

189 HM Treasury. (2007). The Magenta Book: guidance notes for policy evaluation and analysis. Background Paper 7: Why do social experiments? Government Social Research Unit.

190 Tegmark, M. (2017). Life 3.0: Being human in the age of artificial intelligence. Vintage.

191 Norvig, P., & Russell, S. (2010). Artificial Intelligence: A modern approach. Pearson.

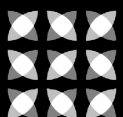
192 Mahesh, B. (2019). Machine learning algorithms - A review. International Journal of Science and Research. https://doi.org/10.21275/ART20203995

193 Australian Evaluation Society. (2024, 19 April). Artificial Intelligence (AI) 101 for evaluators [Video]. Youtube. https://youtu.be/dJrH-ra4VY4

194 Ibid.

195 Thornton, I. (2023). A special delivery by a fork: Where does artificial intelligence come from? New Directions in Evaluation, 178-179, 23-32.

196 Norvig, P., & Russell, S. (2010). Artificial Intelligence: A modern approach. Pearson.



CENTRE
for SOCIAL
IMPACT

www.csi.edu.au