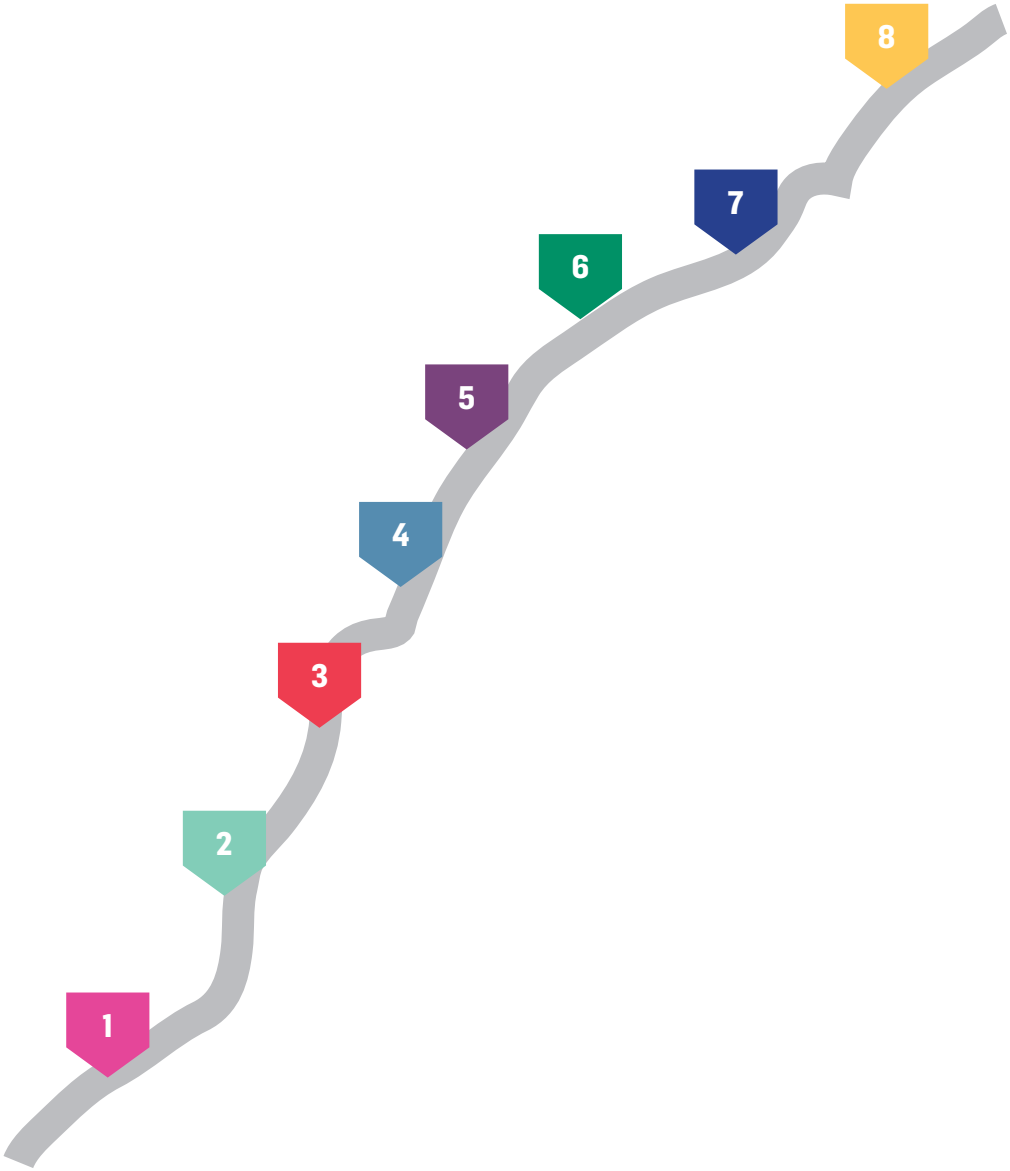


ROADMAP TO SOCIAL IMPACT

Your step-by-step guide to planning, measuring and communicating social impact

CENTRE for SOCIAL IMPACT

UNSW SYDNEY

CONTENTS

Suggested citation:

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.

CONTEXT AND PURPOSE

Roadmap to social impact: your step-by-step guide to planning, measuring and communicating social impact (The Roadmap) is a step-by-step guide to support you and your organisation through the process of outcomes measurement and evaluation, to help you demonstrate your contribution to social impact. We use the term 'program' throughout *The Roadmap* for simplicity, but you may be implementing an initiative, or a suite of programs. If you are implementing actions for a social purpose and want to measure the outcomes of these actions, whatever you call them, this guide is for you.

Each chapter in this guide discusses key concepts, their importance to outcomes measurement and impact assessment, and what they mean in practice.

If you have read CSI's resource *The Compass: your guide to social impact measurement (https://www.csi.edu.au/research/tools-and-guides/compass-your-guide-social-impact-measurement/)*, you will understand why outcomes measurement is important, but you may still have questions about how to implement it. *The Roadmap* is here to help.

WHO THIS GUIDE IS FOR

The Roadmap is for everyone working towards the creation of positive social impact in Australia who wants to measure the change they make for individuals, organisations, communities and the society. 'You' (the reader) may be the person implementing the program, initiative, or actions, the project manager, leader or the whole team. As you will see throughout this guide, the role of measuring is scattered throughout the organisation.

WHAT THIS GUIDE IS NOT

If you want a beginners' guide to outcomes measurement, to understand what it is and why it is important, please read *The Compass*.

While *The Roadmap* introduces some established techniques for data collection and analysis, it is not a guide to research methods. A list of resources is however provided at the end of the guide, including tools for stakeholder engagement, developing a survey or interview questionnaire and data analysis.

INTRODUCTION

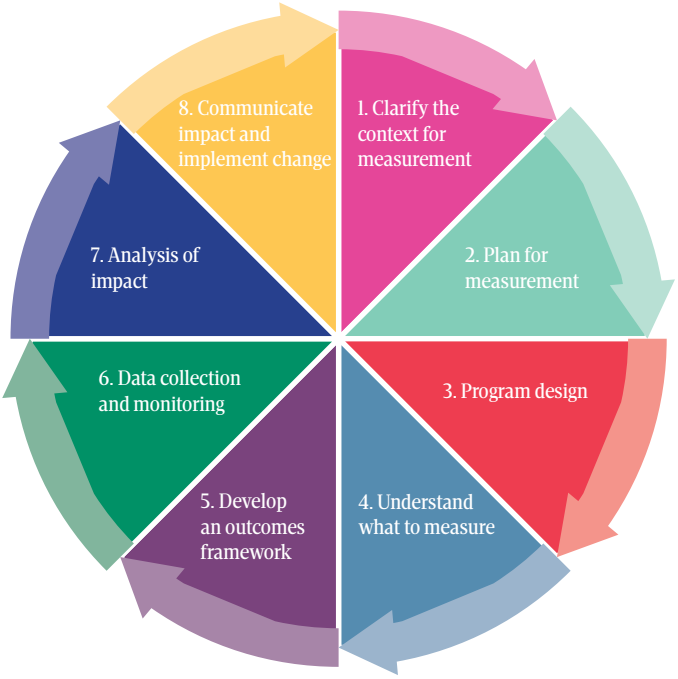
Outcomes measurement is increasingly an integrated part of social purpose programs[i], but it is only effective as part of a wider, ongoing process. *The Roadmap* will guide your outcomes measurement journey through eight steps listed below and in Figure 1 on page five.

1	**CLARIFY THE CONTEXT FOR MEASUREMENT**	a. *Understand your problem* b. *Know your vision, mission, goals and objectives* c. *Recognise whether and when you need to measure outcomes*
2	**PLAN FOR MEASUREMENT**	a. *Understand the need for, and foster a culture of measurement* b. *Know who to engage, and in what capacity (stakeholder analysis)* c. *Unlock your resources*
3	**PROGRAM DESIGN**	a. *Establish your theory of change* b. *Develop a logic model – inputs, activities, outputs, and outcomes*
4	**UNDERSTAND WHAT TO MEASURE**	a. *Develop evaluation questions* b. *Decide the type of evaluation you will conduct*
5	**DEVELOP AN OUTCOMES FRAMEWORK**	a. *Prioritise outcomes for measurement* b. *Identify indicators to measure outcomes* c. *Identify data sources to quantify indicators*
6	**DATA COLLECTION AND MONITORING**	a. *Quantitative method designs* b. *Qualitative method designs* c. *Responsibilities for data collection and monitoring* d. *Ethics and politics of data collection and outcomes measurement*
7	**ANALYSIS OF IMPACT**	a. *Assessing change and impact* b. *Skills and competencies for evaluation*
8	**COMMUNICATE IMPACT AND IMPLEMENT CHANGE**	a. *Sharing your impact with stakeholders* b. *Using results for increased social impact*

Throughout these chapters you will complete 10 core activities.

[i] Throughout this guide we refer to outcomes measurement for 'programs'. The things we talk about can also apply to interventions, organisations and initiatives. You may be implementing a set of actions for a social purpose or social good and want to measure the outcomes of these actions. This guide will support your measurement of all such actions.

FIGURE 1 Stages of Impact Assessment



THE LANGUAGE OF IMPACT ASSESSMENT

The language of impact can be confusing at times. The box below provides definitions for some of the terms you might have come across in the past, such as outcomes measurement, outcomes evaluation or social impact assessment.

PRACTICE YOUR SKILLS

A case study is used throughout the guide to demonstrate how the concepts described work in practice. **Sport** is a fictional program providing free after school sport activities for primary school children living in low socioeconomic status areas. The purpose of the program is to increase physical activity among children, providing them with the opportunity to exercise two to three times a week while practicing various team and individual sports on the grounds of their school. In addition, the program provides information sessions and printed resources for students, teachers and parents about the benefits of an active lifestyle alongside other healthy habits such as good sleep, nutritious food and outdoors time. The program is free of charge to schools and students, is funded by the local government and delivered by a not-for-profit with the support of the Department of Education and Department of Health.

Each section of the guide explains how impact assessment concepts apply to the **Sport** case study. Activities and/or reflection points at the end of each chapter will help you establish and practice the skills to complete outcomes measurement for your program.

The language of impact assessment

- **Evaluation:** An objective process of understanding how a program, policy or other intervention was implemented, what effects it had, for whom, how and why[1]. In an evaluation, social research procedures are systematically applied to assess the conceptualisation, design, implementation, and utility of programs or interventions.
- **Outcomes evaluation:** The assessment of the changes resulting from the implementation of a program, policy or other intervention. It includes both intended and unintended outcomes for a range of stakeholders engaging in a program or intervention.
- **Process evaluation:** The investigation of the extent to which a program or intervention was implemented as planned. It helps understand why changes occurred.
- **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.
- **Outcomes measurement:** A systematic way to assess the extent to which a program has achieved its intended results.[2]
- **Social impact:** The intended and unintended social consequences, positive and negative, of programs (interventions, policies, plans, projects) and any social change processes invoked by these.[3]
- **Social impact assessment:** The processes of analysing, monitoring and managing social impact.[3]
- **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.

1. CLARIFY THE CONTEXT FOR MEASUREMENT

Outcomes measurement is most effective the earlier you start to think about and plan for it. Ideally, this means from program design stage, however, it is better late than never! To understand what you need to measure, you need to recognise ***what*** problem your program is trying to solve, ***how*** it will resolve that problem and with what resources[5]. This guide will provide you with an approach to understand the problem, including its causes and effects. This is the first step in identifying the outcomes your program seeks to achieve.

UNDERSTAND YOUR PROBLEM AND THE SYSTEM IN WHICH IT EXISTS

What problem are you trying to resolve?

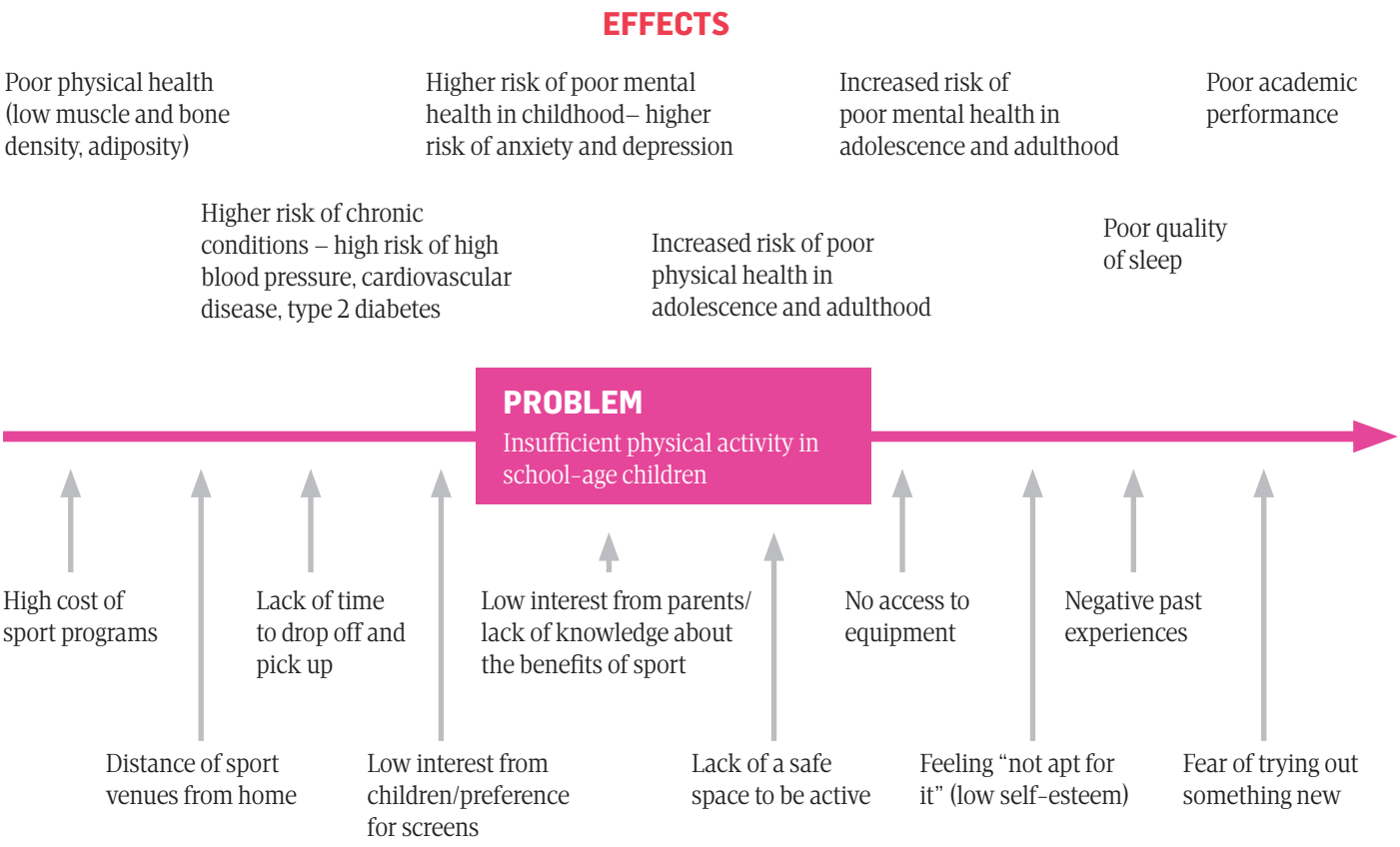
Social problems are often complex, or wicked[6], have a range of causes and effects and often need the effort of multiple programs to be resolved. **Problem analysis** helps understand the entrenched nature of social issues, identify the 'root causes' and helps map potential interventions.[7] This can help you identify the extent of the cause your program is addressing, and which effects it might be reducing. It can also help you identify potential partners or alternative programs that (should) work alongside your program to address the complex problem. While your colleagues and stakeholders are an invaluable source of knowledge to develop this analysis, both causes and effects should be evidence-based, meaning they should be based on existing research and literature on the topic[ii], knowledge and expertise. For **Sport**, our fictional example, the problem that the program is addressing is 'insufficient physical activity among school-aged children'. Indeed, this is one of the elements contributing to the broader problem – unhealthy lifestyle, alongside inappropriate diet, insufficient sleep or extended screen time[iii]. While insufficient physical activity is part of a larger problem, **Sport** is only looking to address the issue of insufficient physical activity among school-aged children.

The metaphor of a tree (see Figure 2) helps identify and visualise the root causes of the problem our case study program **Sport** seeks to address and the effects that it will alleviate or eliminate. Your program or organisation may be looking to address some of the causes and may have the capacity to alleviate some of the effects of the problem you identified. Analysing the full problem holistically with help you understand the space where you operate, the part of the problem you are addressing, give you a first indication of potential outcomes you will be expecting (alleviation of which effects) and may help you identify the need for partnerships to help you address elements of the problem you cannot address on your own. It might not be always easy to clearly map all causes and effects, so looking at the big picture, or the whole system in which your program and the problem exist through systems thinking will help you. For example, a cause for not participating in outside-school sport activities may be a lack of interest in such activities, but other causes may relate to the wider system, for example lack of sport venues near home, no transport options to travel to available venues, or parents being engaged in work at the time when sport activities are available.

[ii] You may want to start this literature review with academic publications that have researched your topic, or published information from similar programs or the industry, both in Australia and overseas.

[iii] Australian Institute of Health and Welfare, 2020.

FIGURE 2 Problem Analysis – Sport



Sources: John Hopkins Medicine, nd; Molnar et al (2004), Australian Institute of Health and Welfare (2020); Committee on Physical Activity and Physical Education in the School Environment, Food and Nutrition Board, Institute of Medicine (2013).

Note: This problem tree is for illustrative purposes and may not include all causes and effects related to youth mental illness. Similarly, not all causes and effects may apply to all who experience mental illness during youth.

What is the big picture or the system?

Social problems are not isolated, they exist within systems. At this point you should think about the wider system in which a problem exists. You will need to consider the various groups or stakeholders[iv] who exist in the system and how they relate to the problem, your program and to each other. Figure 3 will help you visualize potential stakeholders or elements of the system you or your beneficiaries may be engaging with, the nature of the interaction, your levers for change.

You should start to think of the problem from the perspective of the beneficiary and understand how the various layers of the system affect them. For example, some elements of the system for our case study are: the student, their family and home environment, the school, services available, accessibility, past experiences. These elements interact and reinforce each other while presenting 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.

[iv] Stakeholder analysis is further discussed in the next section, from the perspective of stakeholders engaging in measurement.

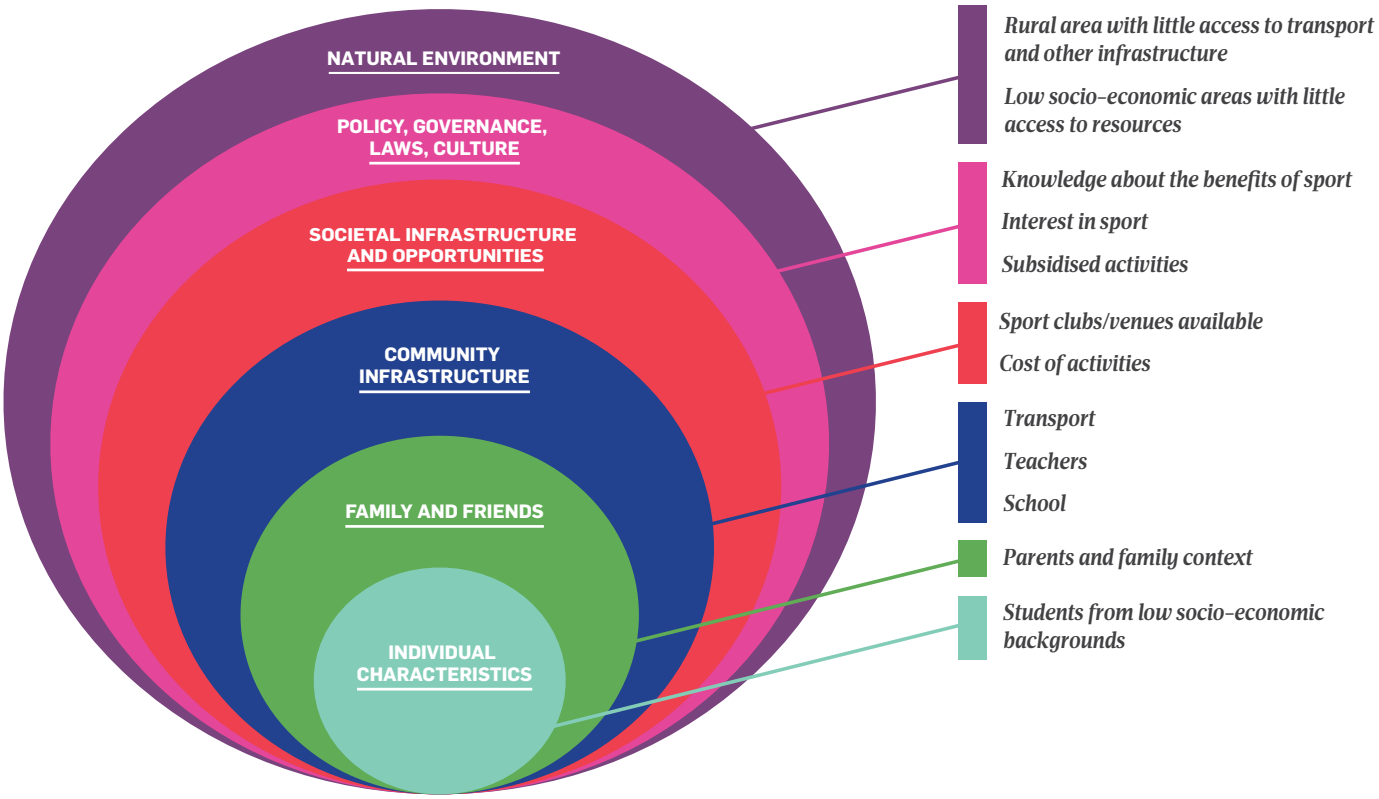
Systems and systems thinking

A system, human-made or natural, is an interconnected set of elements that is coherently organized in a way that achieves something[50]. 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 influencers at the individual, household, community, infrastructural, political and societal level[8] and the stakeholders behind these influencers. Working with these stakeholders to take actions will create change in the system, sometimes in unexpected ways, as the system adapts to change.[9] A strong systems-approach identifies potential intended and unintended consequences, groups that engage and interact with the program and each other, informing you on which levers should be pulled, or not. Thinking about the system will help you understand the context of your program and the problem you seek to resolve. When one element of a system changes, the other parts will be affected and, in the end, the stability of the whole system. Systems thinkers use a few guidelines, or have a few habits:[51]

- 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 include feedback loops (Feedback loops show at least two factors relate to each other in a circular fashion. When one factor changes, how does it impact another related factor?) and have a causal map to show the multiple relationships within the system, between actions and effects.

FIGURE 3 Impact Assessment System and Context, Sport Example



Note: Adapted from Dass et al (2020) (53).

ACTIVITY #1: PROBLEM ANALYSIS

Develop a problem tree for your program/organisation

Supporting activity: start with a mapping of the system to understand the problem in a wholistic way and engage with evidence (literature and practice), stakeholders and practitioners to map the potential causes and effects of the problem. Imagine your beneficiary at the centre of the system and the elements they interact with at micro, meso and macro levels – family and friends, community, infrastructure, societal infrastructure and opportunities, policy, natural environment (e.g. the young person, their family and friends, their school and teachers, the services available to them, access to programs and support, the wider community). Map how these elements connect, reinforce each other or what can hinder your beneficiary. What are your levers for change?

WHAT ARE YOUR VISION, PURPOSE, MISSION, GOALS AND OBJECTIVES?

It is important to align your program and program objectives to the **vision**, mission, purpose and goals of your organisation[10]. The vision is an organisation's statement of its overall ideal and the ultimate goal of its operation[11]. It describes what the future should look like. The **mission** describes 'the business' of the organisation[12] 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 and serve its purpose through several programs or interventions. The vision will provide strategic direction and facilitate decision-making, while the mission will ensure your activities align with the overall purpose of the organisation. The **purpose** is why an organisation exists. Some organisations have shifted in the past years to formulate a purpose statement rather than a vision. In a nutshell, vision, mission, and purpose answer the following questions:

- **Why** do you exist? (Purpose)
- **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, purpose, mission, goals and objectives should be well aligned with the problem you are looking to resolve. Figure 4 presents the vision, purpose, 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[13].

FIGURE 4 Vision, Mission, Goals, Objectives, Sport

Sport vision, mission, values and goals

Vision: Healthy children, healthy adolescents, healthy adults.

Purpose: Ensure school-aged children maintain healthy levels of physical activity.

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

Goals: Reduce life-style 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.

ACTIVITY #2: CLARIFY OR (RE)FORMULATE YOUR ORGANISATION/ PROGRAM'S VISION, PURPOSE, MISSION, GOALS AND OBJECTIVES

Do you have a program-level mission, an organisation-level mission? Does the program mission align to the organisation vision?

SHOULD YOU MEASURE OUTCOMES?

Measurement for the sake of measurement can be harmful to programs and progress. It may be that what you intend to measure is not yet measurable (e.g. the outcome has not been achieved yet), or that measurement interferes with program delivery (e.g. data collection may 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'[4,14].

Evaluability assessment tests:

- whether a program is ready for outcomes measurement (and evaluation), or
- when outcomes measurement and evaluation would help improve the program.

Outcome measurement is the first step towards evaluation. Once data to measure outcomes have been collected, it is the role of an evaluator to analyse this data and complete an evaluation of the program . The evaluator can be internal to the program[v] (e.g. a manager or internal researcher) or external. There are advantages and disadvantages to having an internal or external evaluator, relating to cost, knowledge, flexibility, objectivity, accountability, willingness to criticise, ethics and utilisation of results[15].

The evaluator will 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 evaluation from the beginning and build in data collection time to ensure the evaluation is reliable and achievable. Your program may not be ready to be evaluated but having an outcomes measurement plan will ensure evaluation is achievable down the track.

Evaluability assessment involves a six-step process[16,17]

- Involve key stakeholders (e.g. policymakers, managers, staff) – to ensure the program theory conforms with their expectations (stakeholder analysis, Section 2).
- Clarify program design – ensure the relationship between inputs, activities, outputs and outcomes is as expected from the points of view of key policy makers, managers and interest groups (logic model, Section 4).
- Clarify program reality – whether the program was/is implemented according to the program design (logic model, Section 4).
- Assess the likelihood that the program activities will lead to the intended outputs and outcomes (logic model, Section 4).
- Agree on required changes to the program design (implementation, Section 7).
- Agree about the intended use and value of future evaluation activity (communication, Section 7).

[v] Types of evaluation and associated analyses are discussed in Section 6.

2. PLAN FOR MEASUREMENT

Overall, outcomes measurement is beneficial to organisations for several reasons[vi]. Before diving into measurement, you need to ensure that your organisation's strategy, culture, engagement and human resources are set-up (or build them!) to support outcomes measurement. Your organisation should have an established **culture of measurement** and understand the **importance and use of outcomes measurement for all stakeholders**.

FOSTERING A CULTURE OF MEASUREMENT

Outcome measurement does not happen in a vacuum; it requires an organisation that is ready, willing, and able. An organisation with a strong measurement culture engages in self-evaluation, self-reflection, and self-examination[14]. It considers the impact it is seeking to achieve, takes responsibility for it and actions results to challenge or support its activities[18]. It values candor, challenge and genuine dialogue, with staff able to use the language of measurement. A strong measurement culture supports experimentation and risk-taking and learns from mistakes and weak performance[14]. Outcomes and impact measurement are visible on meeting agendas, in annual reports, on the website, and in performance reviews. The leadership team lead by example, building capacity for, and investing in measurement, while being held accountable for results and measurement culture.

Measuring outcomes provides

- Accurate judgement about the value of a program
- An evidence base on program effectiveness
- Accountability and efficiency: a critical tool for resource allocation decisions
- The basis for learning and responsible policy development within organisations
- The key ingredient for evaluation, strategic planning and good governance
- Staff engagement and motivation
- Data required by, and to attract, funders.

How to build a measurement culture

Understand your organisation's position: Self-assessment tools (see Appendix 2) can help assess the extent to which outcomes measurement is embedded in your organisation, inform an action plan, and monitor progress.[9,14] Used across the organisation such self-assessment programs can be a conversation starter, an early process in engagement.

Leadership: A guiding coalition of champions, participants, influencers, change agents, and communicators lead strong measurement culture. The Board, CEO, and Executive should be champions and provide structure including incentive systems, clear roles and responsibilities, performance review, and reporting mechanisms. Their own reporting and accountability should be results-led. Assess the skill set of your Board: ensure there is someone with measurement expertise who will inform demand for results-based information, and ask critical questions[14].

Systems: Assess your current policy, procedures, data management systems, and accountability plans to see if they align with and support outcomes measurement. Does infrastructure (such as IT platforms) need to be developed? Is program documentation in order? Are there ways to integrate with existing data collection and reporting systems? What resources will be required?

[vi] See The Compass[6] for this discussion.

Capacity, capability, and connection: What 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, service networks or peak bodies), partnerships, mentors, and universities (academics, students).

Learning orientation: Outcome measurement is ultimately about learning and action. Your organisation should build opportunities for learning through communication loops, regular discussion (such as at team meetings), training, mentoring, and conferences. Results need to be mined for what they reveal is and is not working. This learning needs to be acted on by stopping, growing, or embedding particular approaches. Who will decide which action is to be taken? How and by who will this action be monitored? The outcomes and impact measurement loop is cyclical and ongoing!

Understand what merit and quality look like for your outcomes measurement system. **Quality** means the outcomes measurement system connects with your organisation's mission and values, and will include integrity, respect, responsiveness (adaptation based on results), stakeholder involvement, transparency in communication, and being culturally responsive. **Merit** means:

- Applying established and appropriate methods
- Focusing on all the types of impact created (positive, negative, un/intended)
- Attribution (claiming only the difference you know you've made)
- Utility (application)[19].

? *Reflection*

1. Which of the above steps could you implement in the short-term? Which are long-term prospects?
2. What will you commit to do in the short-, medium- and long-term? What could you do tomorrow?
3. What strengths can you leverage?
4. An out of the box idea is....

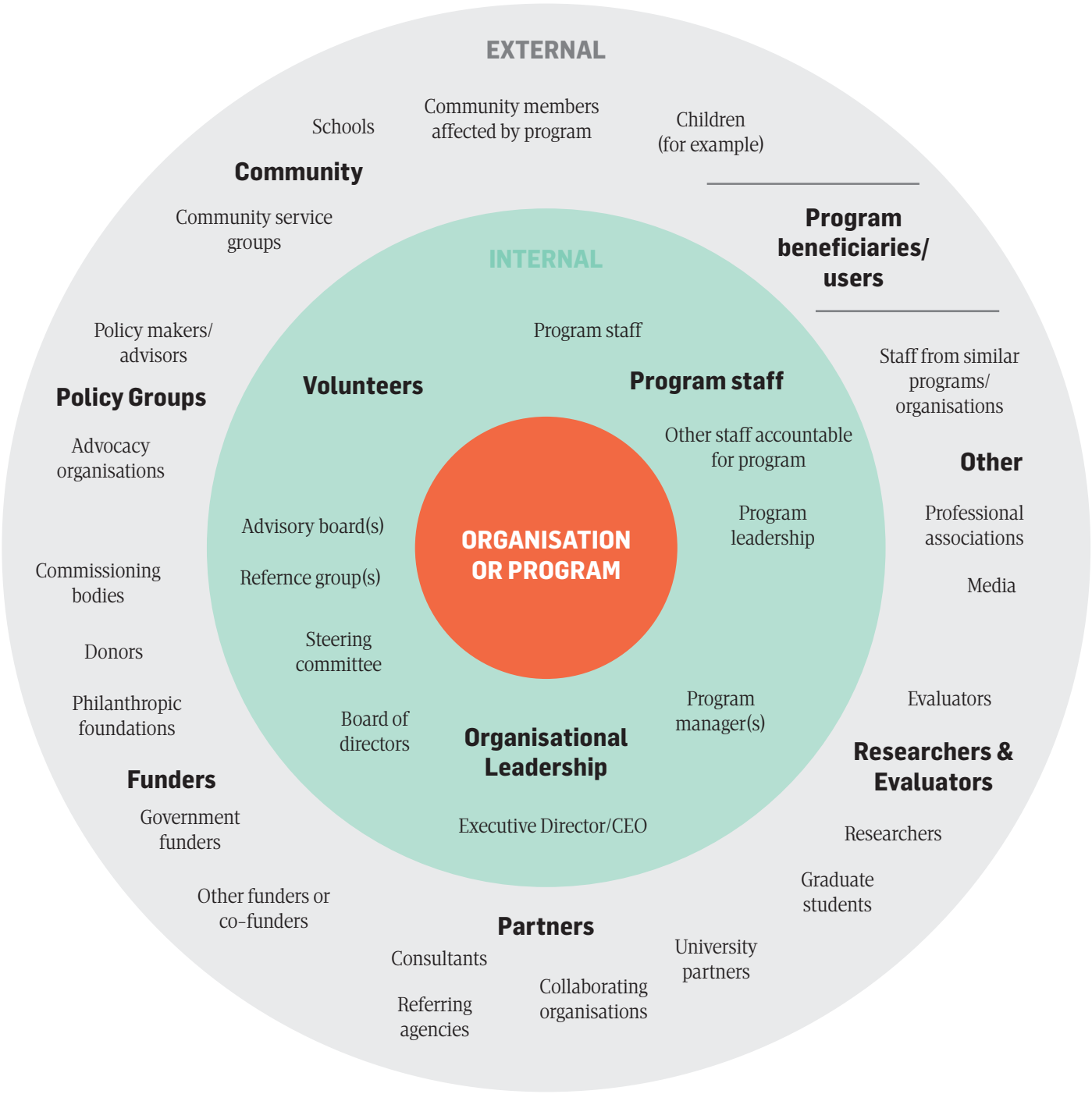
KNOW YOUR PEOPLE: STAKEHOLDER ANALYSIS

Outcomes measurement and impact assessment are more likely to be relevant, thorough, actioned, participated in, of good quality, and successful if your stakeholders are engaged with the process. The first step in achieving this is understanding who your stakeholders are, their current and potential level of engagement with the program, and their attitudes and aptitudes for measurement.

Identifying and analysing your stakeholders

The diagram below (Figure 5) provides an example list of internal and external stakeholders who might be relevant to your organisation or program. While in the previous section you read about understanding stakeholder within the system from micro to meso and macro levels as they relate to the individual, this provides you with a different approach to understand your stakeholders, in terms of their positioning relative to the program or organisation.

FIGURE 5 Stakeholder Groups Relative to the Organisation or Program



Alongside their roles in the program, the priorities, interests, and needs of your stakeholders for measurement need to be understood. Consider ***what they bring to measurement***, ***how important*** their perspective is, and ***what may motivate them*** to participate[20]. Some stakeholders may seem peripheral, yet important to engage. Identify which stakeholders might have ***resistance*** to what you are trying to achieve (and how to address their concerns), how to increase ***engagement*** (and how to sustain it) and who might be ***champions*** (and how to empower them). Outcomes measurement may be met with resistance due to a lack of internal capacity, especially within smaller organisations, to measure outcomes, lack of funding, a perceived feeling of knowing 'I am doing good' hence no need to measure, or that clients wouldn't care if they measured.[21]

For example, in the **Sport** case study, there might be resistance to measurement from staff implementing the program as they have low skills in data collection and find it a burden. Yet, they may become champions if they are engaged in measurement early on, trained and provided with the tools to measure, 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[22]:

- 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.

See also Figure 19 in Appendix 3 for a further description of these types of stakeholder engagement.

Think about the parts of the measurement process with which your stakeholders will be involved: planning, design, question development, data collection, review, action plans. Think also about the control they have over these processes. Social impact takes place in a political context. The political context is especially important to understand in social impact assessment as this often focuses on the reallocation of resources, serves vulnerable groups, and engages a range of stakeholders with complex relationships[40]. Your stakeholder analysis should support your understanding of relationships and politics surrounding your program.

Issues of budget, geographic location, ensuring diverse perspectives, decision-making processes, exit strategies, stakeholder capacity and measurement capability, and organisational capacity for stakeholder engagement, all need to be considered as part of your stakeholder engagement strategy.

Putting Users at the Centre

As program beneficiaries usually represent a high priority stakeholder group, it is good practice to consider them as central to your measurement process and decision-making. This reminds us that measurement is 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. It fosters a sense of inclusion, agency, and contribution. And improves your measurement process by ensuring meaningful measures, completeness and acceptability of tools, and broadening dissemination[23].

Mechanisms for engaging your community in measurement include reference and advisory group membership, champions, providing expert review and development of tools, and co-design of methods and communications.

While there are some challenges to engaging users, there are assumptions about involvement that deserve disrupting. Challenges might include unequal power relationships, representation, resourcing, thinking 'it's too hard', and assumptions about whether consumers are willing and able. Organisations need to be willing to change their structures and communications, as well as provide support and training to consumers, to facilitate meaningful participation. See the range of tools to facilitate stakeholder engagement in Appendix 2.

Understanding who the stakeholders are, how they interact with each other and the program, and their attitude and need for measurement will not only support the delivery of the program but the data collection, outcomes measurement and evaluation.

ACTIVITY #3: MAP YOUR STAKEHOLDERS

Map your stakeholders and their engagement in the program

UNLOCK YOUR RESOURCES

At the beginning of a program, it can be hard to know the resources you will need to measure your outcomes. For this reason, you might need to come back to this step after you developed a good understanding of the evaluation type you need to implement and the data you need to collect and analyse (Section 7). You first need to understand what data you will need to collect, the frequency of data collection, number 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. Consider:

- 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).
- Alternative sources: administrative and secondary data, other organisational data readily available.
- Who could collect data and when (Could some additional information be collected at in-take?).
- Should data monitoring and analysis be done internally or externally (Would training be cost saving in the long-term?).

You also need to plan for resources: allow time for staff to train in data collection and monitoring, time for the actual data collection, time and skills for data analysis (which will vary with the type of analysis 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[24]. You should base your cost estimate on previous experiences, expert advice, and thorough planning.

The risk of under budgeting for outcomes measurement is high, including inability to capture all outcomes and misrepresent program achievements. Not allocating sufficient resources (staff, time and money) to communicating findings can make outcomes measurement redundant, through missing out on the opportunity to engage relevant stakeholders and implement change[45]. There is a range of free resources to support organisations looking to complete outcomes measurement, some of these are listed in Appendix 2.

Some types of measurement are more expensive than others and may need expert advice. Considering the need for resources from the beginning will help you ensure you are setting realistic goals for data collection and analysis. If, given your available funding, outcomes measurement and evaluation are restricted, you might need to look for funding alternatives.

? *Reflection Points*

- As you proceed through the next steps of this guide, consider the resources you will need for:
  - » Program planning
  - » Outcomes measurement planning
  - » Data collection
  - » Data analysis and evaluation
  - » Report writing
  - » Dissemination of findings
- What do you need to do to unlock the resources you require for outcomes measurement?

3. PROGRAM DESIGN

WHAT WILL CHANGE: THEORY OF CHANGE

A theory of change is an explicit theory or model of how a program will achieve the intended or observed outcomes.[25] 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.[26] 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 your organisation to understand how your program will achieve its goals. It will help you with[27,28]:

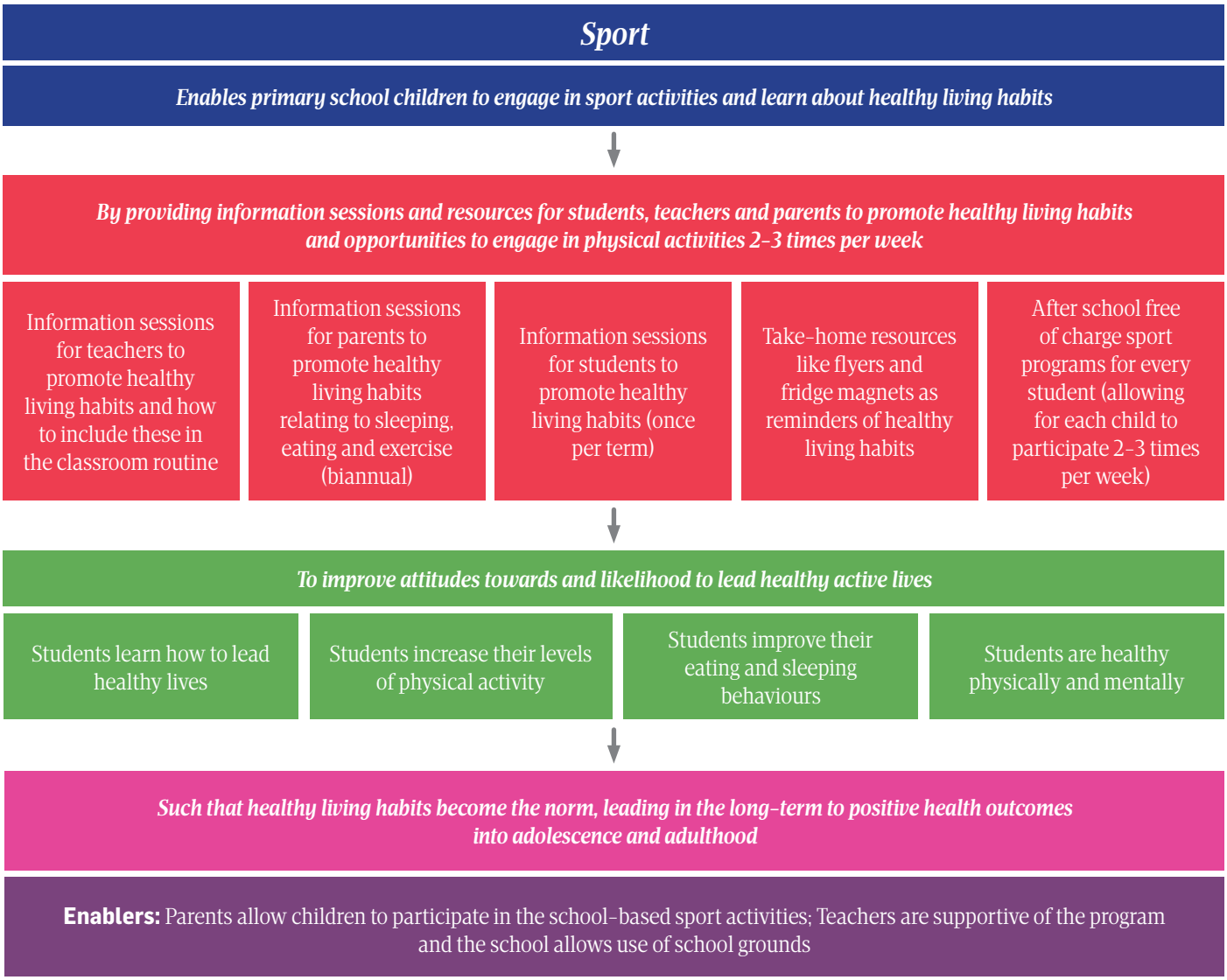
- **Strategy**: Helps teams work together to achieve a shared understanding of a program and its aims; ensures all activities align with the purpose of the program; encourages in-depth thinking about the program and its assumptions.
- **Measurement**: Helps to formulate and prioritise evaluation questions and plan evaluations; 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 knowledge 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[vii].

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. 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 6 for the theory of change for **Sport**.

[vii] Some theories of change also discuss assumptions (why activities lead to certain outcomes, or why an intermediate outcome leads to a long-term outcome), but we include these in the logic model, which further details the theory of change.

FIGURE 6 Theory Of Change, Sport



ACTIVITY #4: THEORY OF CHANGE

Develop/review a theory of change for your program

Tips on developing a theory of change

Start with your team and consider including other key stakeholders. Organise a workshop and prepare flip-chart paper, post-it notes and texters. Or use an online tool where everyone can contribute, like Google Docs. Explain the components of a theory of change (activities, long-term outcomes/goals, enablers). 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. 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. Take time to discuss, remove duplicate ideas/concepts, rearrange for timeline, 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 logic model when you expand on the theory of change. When you are confident with the draft theory of change, circulate it to other stakeholders and ask for feedback. Remember the 'Ikea effect'[29] – people relate more and have a greater commitment and ownership to things they helped to create!

MAP YOUR PROGRAM: LOGIC MODEL

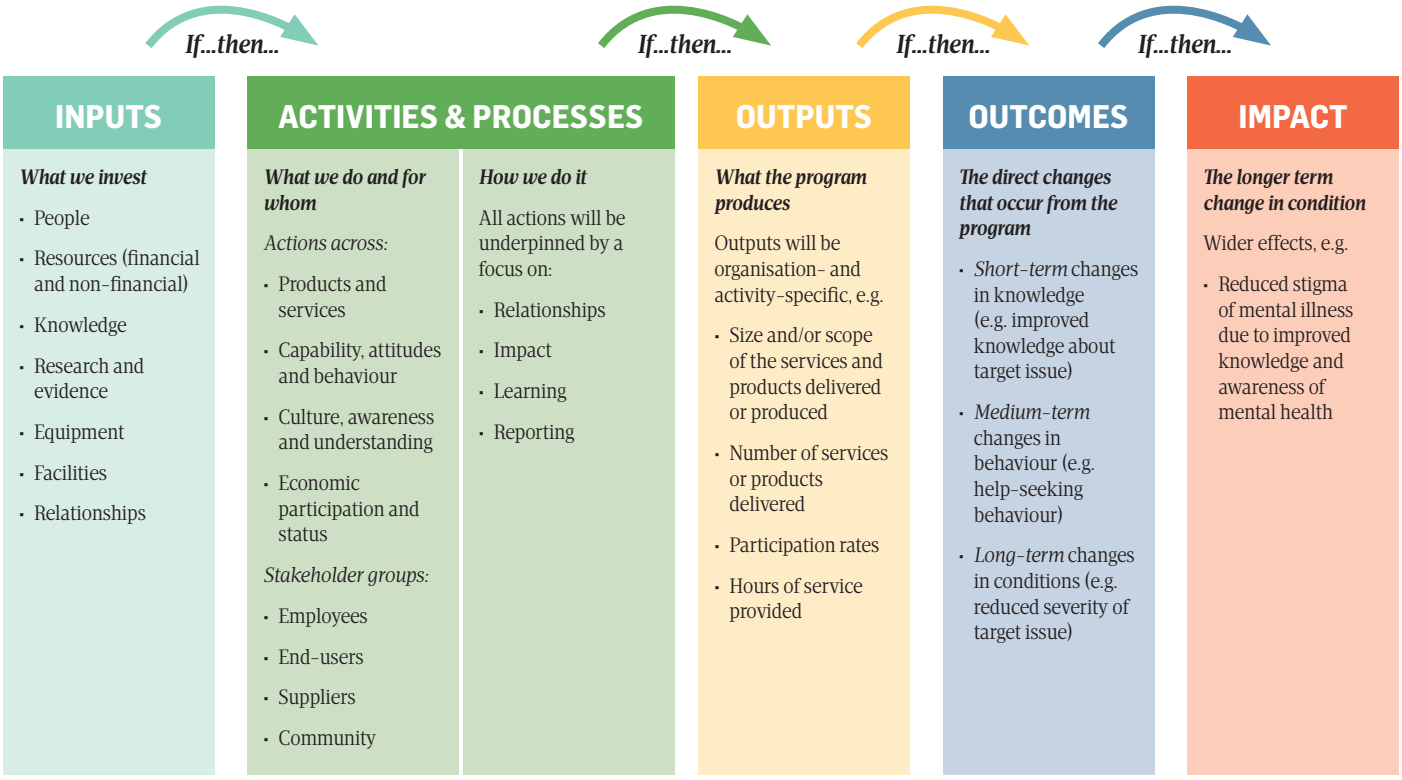
A logic model is a visual representation of how your program will achieve its goals, including the short-, medium- and long-term outcomes (discussed below, see also SECO logframe matrix[7], and Muir and Bennet 2014[6]). Like your theory of change, your logic model is best developed at the design or planning stage of a program, but if this has not happened, these can be developed, modified and enhanced as the program evolves. Use evidence to link activities to outputs and outcomes remember that outcomes are based on assumptions (e.g. we assume that if students are offered the opportunity to participate in organised sport activities after school they will participate 2-3 times per week and their physical health will improve).

The logic model (Figure 7) has an underlying "if-then" relationship, linking a program's necessary **inputs**, **activities**, **outputs**, **outcomes** and **impact**. Assumptions and risks will accompany your logic model: these are external conditions which could affect the program's progress, but which are not under the direct control of people implementing, managing or planning 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.[7] 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 into 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 realization of outcomes (risk). Or the risks might be at the school-level, for example lack of infrastructure to support the proposed sport activities.

Logic model terms

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

FIGURE 7 Logic Model Template



Developing the first half of your logic model – 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 deliver of these activities (how much of these activities the program will deliver, how many clients it will engage with). Figure 8 presents the inputs, activities and outputs for **Sport**.

FIGURE 8 Logic Model Sport (Part 1)

Inputs	Activities	Outputs
• Staff (trained PE staff to deliver the programs) • Educational materials • Funding • Partners (health and sport organisations) • Infrastructure (school grounds available; transport for staff to access schools) • Equipment	• Prepare and distribute resources to schools • Deliver information sessions to teachers, parents, students • Deliver after school daily sport sessions	• Number of schools taking up the program • Number of information sessions delivered to teachers, parents, teachers in one year • Number of participants to each information session • Number of students enrolling in sport activities

Note: The inputs, activities and outputs are for illustrative purpose and may not be the complete collection for the case study.

The second part of the logic model, mapping the outcomes of the program, can be more challenging due to difficulty in identifying outcomes or confusion between outputs and outcomes.

OUTCOMES

Outcomes – what a program achieves – can be measured at different points in time[viii] and at different levels.[31] **Short-term outcomes** capture changes in knowledge (e.g. improved knowledge about benefits of regular exercise), **medium-term outcomes** capture changes in behaviour (e.g. engagement in regular exercise) and **long-term outcomes** capture changes in conditions (e.g. reduced rates of obesity among school-age children, adolescents and adults). There are no definitive guidelines on the timeline 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 type of the problem that the program is addressing, the purpose, scope or the target population.

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.

Some evaluation techniques, such as Social Return on Investment (SROI), rank outcomes in terms of their importance to stakeholders, but this is not common practice in non-financial valuation techniques such as logic models or outcomes evaluations[ix]. In the context of impact investment, identifying a single 'primary' outcome will guide the size calculation and basis of payments in a social impact investment, accompanied by secondary outcomes that complement the primary outcome.[2] Terminology should not interfere with the value of the full suite of outcomes in any program, that is, a secondary outcome should not be considered less important than a primary outcome.

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 **Sport** can measure their long-term impact (improved health outcomes in adolescence and adulthood) but they can measure the change in the levels of physical activity (medium-term or intermediate outcomes), which serves as a proxy and may predict the ultimate outcomes.[2] The extent to which intermediate or medium-term outcomes can serve as proxies is not straight-forward and it requires a thorough literature investigation, consultation of organizational data or with experience practitioners. Additional activities 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.

[viii] Sometimes these are referred to as immediate, intermediate and long-term[30]

[ix] See Section 4 Outcomes framework for further discussion on prioritising outcomes.

You should also consider whether your program yields financial, social 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).

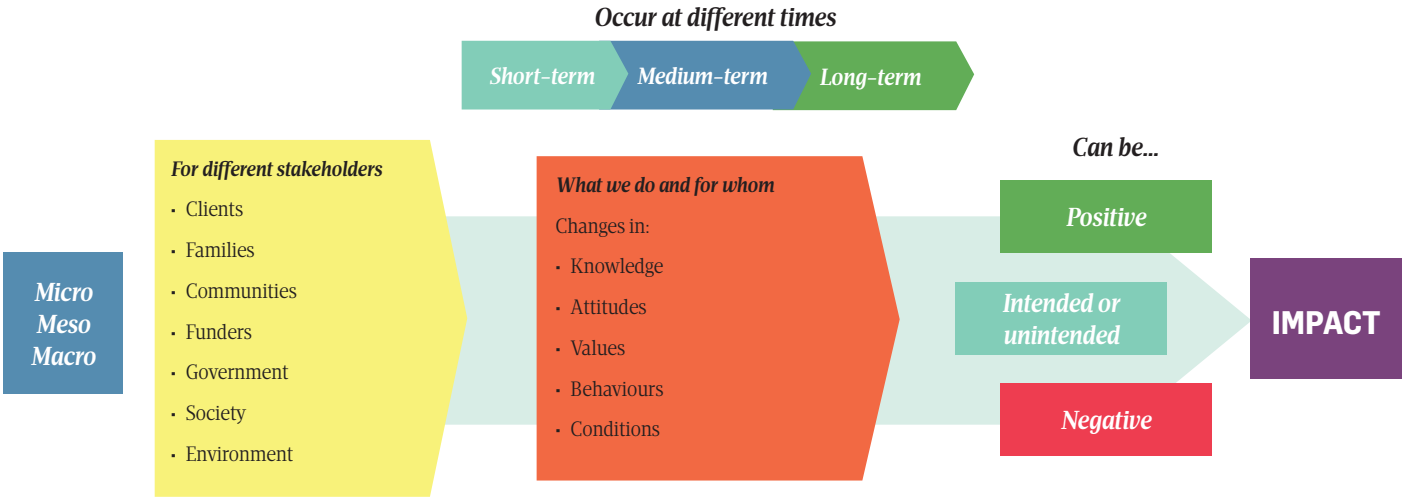
Remember, not all outcomes are predicable. It is often hard to project **unintended outcomes** (positive, negative, or neutral), but these may become obvious as the program matures, and it is important to allow for these to be measured. Collecting qualitative data from a range of stakeholders is a good 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).

Impact is the systemic-level change your program intends to achieve. This relates to the vision of your organisation.

Outcome statements

who/what will change; change/ desired effect (action verb); in what (the expected result); can also include timeline (by when) but can exclude if timeline is clear from the data collection. E.g. young people can recognise signs of depression; young people seek help.

FIGURE 9 Outcome Types



Key points

- Outcomes are not always positive
- Short- or medium-term outcomes 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 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 and environmental
- Intended and unintended

FIGURE 10 Logic Model Sport, Part 2 (Outcomes)

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

Note: The outcomes presented here are only for illustrative purposes and are not a complete map for the program.

ACTIVITY #5: LOGIC MODEL

Develop a logic model for your program[x]

Tips to develop a logic model

Developing your logic model is a good opportunity to engage diverse internal and external stakeholders including the evaluation team, people implementing the program, client representatives, leaders, funders, etc. 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. 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. Shuffle the post-it notes and discuss any points of disagreement or confusion until you have agreed on a logic model 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. Logic model

Theory of change: an explicit theory or model of how a program will achieve the intended or observed outcomes.[25] 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.

Logic model: 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.

[x] Or review, if you already have one

4. UNDERSTAND WHAT TO MEASURE

You have a good picture of the outcomes your program is likely to achieve. It is time to think about how you will measure these. But...

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

Measuring all outcomes may not be feasible due to a range of constraints (resources, time, access to respondents). This is a good time to prioritise the outcomes you will measure. You need to consider your evaluation questions – those are the questions that you want answered. Think again about your 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) and funding (Can you afford it?) available for outcomes measurement. As we need to clarify a few concepts before you can develop the outcomes framework (Section 5), we discuss below the main types of evaluation and evaluation questions.

TYPES OF EVALUATION AND EVALUATION QUESTIONS

Evaluation is an objective process of understanding how a policy or intervention was implemented, what effects it had, for whom, how and why[1]. Well planned and executed evaluation provides evidence for improved design, delivery and outcomes, and supports decision making. Depending on its timing, your evaluation may be[xii]:

- **Formative evaluation:** evaluation with the purpose to improve a model. It takes place during a program's implementation with the aim of improving its design and performance.
- **Summative evaluation:** evaluation with the purpose to judge a model, to assess the extent to which it achieved its intended (and unintended) goals. This type of evaluation happens at the end of a program, or well after a program ended.

'When the cook tastes the soup, that's formative evaluation; when the guest tastes it, that's summative evaluation'.[42]

We have referred throughout this guide to outcomes measurement, implying an evaluation of outcomes. Depending on its purpose, your evaluation may be:

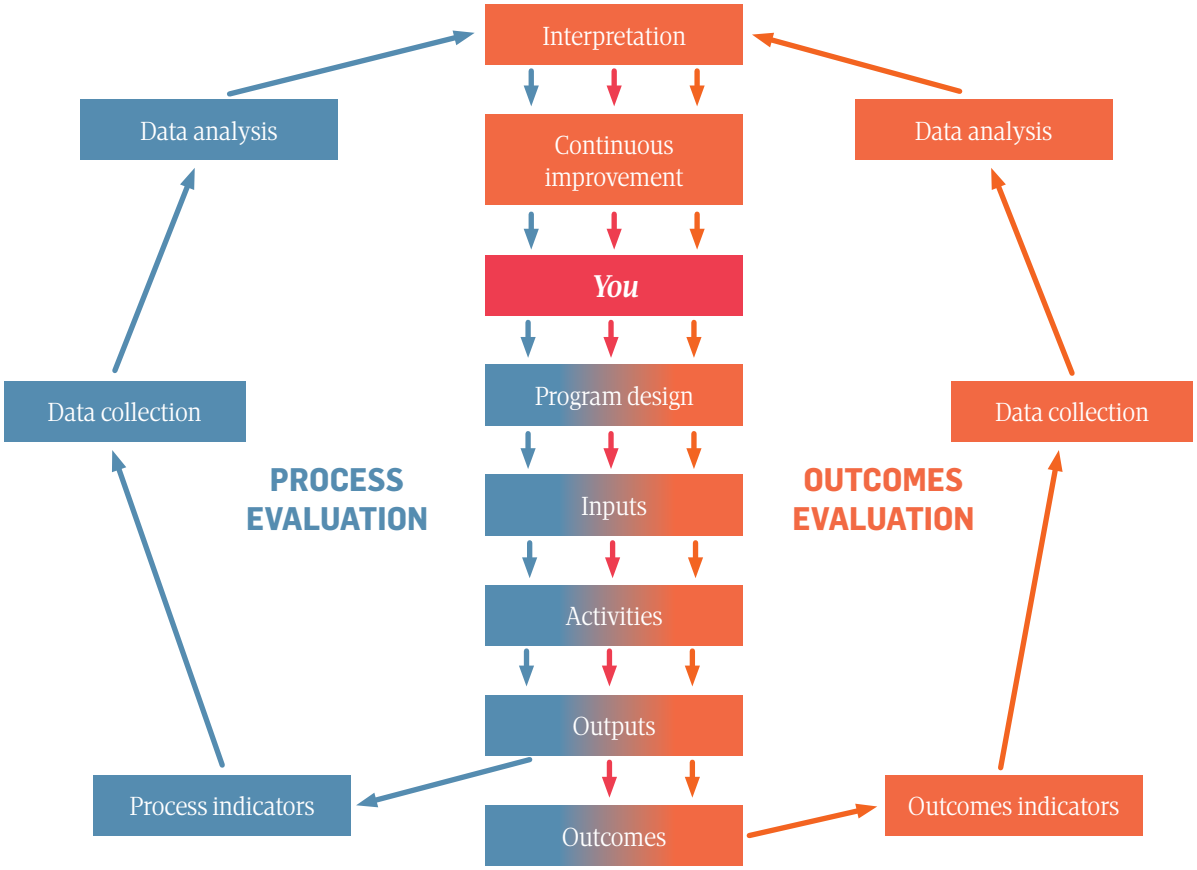
- **Outcomes evaluation:** explores the changes occurring as a result of a program.
- **Process evaluation:** investigates how a program was established and implemented or delivered.
- **Economic evaluation:** studies whether a program generates value for money.

[xi] Quote lined to the work of sociologist William Bruce Cameron.

[xii] Developmental Evaluation is another distinct type of evaluation. This evaluation assists social innovators develop social change initiatives in complex or uncertain environments. It functions much as the formative evaluation, yet there is little knowledge of the outcomes that could or should be achieved, and little understanding of the system in which the program operates. This type of evaluation is most frequent in development projects.

It is important to consider process as well as outcomes evaluation (Figure 11). The former 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 one of the intended activities).

FIGURE 11 Process and Outcomes Evaluation



The Organisation for Economic Cooperation and Development (OECD) developed six evaluation criteria that can also serve as guidelines for selecting evaluation questions:

- **Relevance:** Is the intervention doing the right things? The extent to which the intervention 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 intervention fit? The compatibility of the intervention with other interventions in a country, sector or institution.
- **Effectiveness:** Is the intervention achieving its objectives? The extent to which the intervention achieves, or is expected to achieve, its objectives, and its results, including any differential results across groups.
- **Efficiency:** How well are resources being used? The extent to which the intervention delivers, or is likely to deliver, results in an economic and timely way.
- **Impact:** What difference does the intervention make? The extent to which the intervention 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 intervention continue or are likely to continue.

It may be that not all these criteria are relevant to your program or what you intend to evaluate, but they can guide you into the questions you would like your evaluation to answer. A range of questions that each type of evaluation may answer is provided in Figure 12.

FIGURE 12 Evaluation Questions

FORMATIVE EVALUATION

- What occurred?
- How did it happen and why?
- What design factors were used and how well?
- What new knowledge has been generated?
- How well aligned to existing activity was it?

PROCESS EVALUATION

- Was the method or model adopted?
- Were services offered in scope & to those intended?
- To what extent were quality measures met?
- Did the trial collect valid data and findings?
- Was the activity cost efficient?
- Were the processes used effective in delivery?

SUMMATIVE EVALUATION

- What happened and how well?
- Was performance to the targets set?
- Did people adhere to the standards required?
- Was the program/strategy resource efficient?
- How effective was the effort?

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?

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. Two established approaches for impact evaluation are randomized controlled trials (RCTs) and quasi-experimental designs (QEDs). While there are merits and limitations associated with both approaches, they provide good options to isolate and evidence the impact of an intervention. See Appendix 1 for an introduction to RCT and QED.

FIGURE 13 Sport Evaluation Questions (Example)

Sport evaluation questions:

Were the information sessions delivered as expected?

Did students enroll in the after school sport program?

Did the students improve their physical activity habits?

Did the students change their eating and sleeping habits?

ACTIVITY #6: FORMULATE EVALUATION QUESTIONS

Formulate evaluation questions for your program evaluation.

ACTIVITY #7: IDENTIFY THE EVALUATION TYPE FOR YOUR PROGRAM

5. DEVELOP AN OUTCOMES FRAMEWORK

WHAT IS AN OUTCOMES FRAMEWORK?

Having developed the evaluation questions will give you an indication of the outcomes you want to measure. You can now prioritise the outcomes; look at the logic model you developed, your stakeholder groups, your evaluation questions and flag what you need to measure for the type of evaluation you want to conduct. Selecting the outcomes you need to measure is the first step in developing an outcomes framework.

ACTIVITY #8-1: PRIORITISE OUTCOMES

An **outcomes framework** (also referred to as an 'outcomes hierarchy') is a collection of ***outcomes*** you intend to measure, the ***indicators*** or measures for the outcomes, the ***data sources*** you will use to quantify those indicators, and the ***timing for data collection***. You will learn how to select indicators in this section and Section 6 discusses the various types of data you can collect to quantify indicators.

INDICATORS

Indicators are the measurable markers that show whether change has occurred in an underlying condition or circumstance. Indicators can be expressed as percentages, proportions, numbers, ratios, or perceptions, behaviours, satisfaction, quality. Indicators can be a single measure capturing a condition at a certain point in time, such as the proportion of participants living with a mental health condition, or a composite made up of several measures, such as the Kessler psychological distress scale[32], which measures ten aspects of psychological distress but reports this as one value between 1 and 50.

Various criteria for indicator selection have been developed over the past two decades, including SMART.[33,34] Bennett et al (2016)[7] differentiate further between technical and contextual indicators selection criteria. Technical criteria refer to the extent to which the indicator is a good measure for your outcomes. For example, whether the 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 clients be comfortable to answer certain questions?) or is it feasible (is it practical to collect the respective data?). See Appendix 3 for a description of 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 necessary to measure an 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, both needed to assess change.

STEPS TO DEVELOP INDICATORS:

1. Allow for time and resources to review indicators. Consider: how broad is the review, how long do you have, can you engage stakeholders, 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.
3. Assess indicators against the technical and contextual criteria. 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 all that can be measured.

The language for indicators

Indicators capture at one point in time participation rates, individual behaviours, incidence, prevalence, and attitudes. 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 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. They can also be established measurement scales, like the Rosenberg self-esteem scale (measure of self-esteem), body mass index group, or mental health scales (like Kessler 10, Kessler 6, SF36).

TYPES OF INDICATORS

It is a good idea to assess an outcome through **objective** and **subjective** measures. Indicators are often objective and imply quantifiable concepts measuring how much/many/often. They can also capture subjective responses, such as attitudes and feelings (e.g. changes in quality of life; feelings of anxiety).

'Qualitative indicators' are however a vexed topic, because qualitative data is inherently different to established 'indicator standards' such as validity checks, replicability, and standardisation. It is recommended to collect qualitative data alongside quantitative data to give a sense of what the outcome looks like 'on the ground' when a quantitative indicator is improved for a person, community or population. This data enables a program to 'tell the story' of impact – what it looks or feels like in people's lives. Qualitative work 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 have a high literacy level. See Figure 9 for some examples of quantitative and qualitative measures and the next section (section 5) for insights into data collection techniques for quantitative and qualitative indicators formulated at individual, community or societal level.

FIGURE 14 Outcomes and Indicators, Sport

Outcome	Indicator
Students engage in regular exercise (medium-term)	Proportion students participating in 2-3 after school sport activities per week Proportion of children undertaking at least 60min of physical activity each day
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?

*These outcomes and indicators are for illustrative purposes, they are not an exhaustive list for the fictional Sport program.

ACTIVITY #8-2: DEVELOP INDICATORS TO MEASURE THE OUTCOMES YOU PRIORITISED.

To complete the outcomes framework, you will need to identify data sources to quantify the indicators. These are discussed in the next section of the guide, Section 6. See Appendix 3, for an Outcomes Framework template.

6. DATA COLLECTION AND MONITORING

You developed indicators to measure the outcomes for your program. To complete the outcomes framework, you must decide the most appropriate data collection tools to quantify the indicators. Here we explore different ways to collect and monitor data.

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

Baseline data and benchmarking

Data collected prior to the program is **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.[xiii] 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 DATA DESIGNS

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

Surveys

Surveys are standardised data collection instruments that are usually administered face-to-face, online, by phone or post to generate quantitative data. They may also collect qualitative data, often regarding people's experiences and attitudes[xiv]. 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, skills and likelihood to respond)
- Budget (online surveys are cheaper to administer than post, phone or face-to-face)
- Type of questions (some questions might need visual supports; complex questions may be easier to design in online formats)
- Will respondents be more likely to share accurate information if the interviewer is present or absent?

[xiii] For example, such baseline data could be available in administrative records.

[xiv] 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. See Appendix I for resources with tips in survey design.

Administrative data

Administrative data is program data collected for all participants. For example, data a case worker might record about a client after each encounter, or headline data an organisation might use in annual reporting (e.g. proportion of female clients) is administrative data. While the primary use for this data is administrative rather than research, it is helpful for capturing populations who may not respond to a survey, rich information about the same individual, provide information for potential comparison groups for your evaluation, and to conduct complex statistical analyses due to large sample sizes. Program data is often collected on participant intake 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). Using such data has advantages and disadvantages (See Figure 15).

FIGURE 15 Sources of Quantitative Data

SURVEYS

- Standardised data collection instruments.
- Generate both quantitative and qualitative data – most often with regards to people's experiences and attitudes.
- Reach large numbers of people for a relatively low cost and can be repeated to track changes in people's behaviours.
- Response rates, however, can be low which can jeopardise the validity of data collected.
- Surveys can be administered at program level, organisation or sector level, or at national level. They can be administered face-to-face, online, by phone or by mail.

ADMINISTRATIVE DATA

- Program data collected for all participants.
- On a larger scale, administrative data is data collected for the purposes of registration, transaction and record keeping, usually during the delivery of a service by government departments and other organisations (UK Administrative Data Liason).
- Using this data for research allow noe to capture populations who may not respond to a survey, link various information about the same individuals and identify counterfactual and control groups.
- Program darta is usually collected on a participant intake form and is renewed regularly, as necessary.

SECONDARY DATA

- Sceondary data are collected by someone external to your program e.g. national data sets, or administrative or survey data collected by a different organisation.
- Advantages: low cost, may have large sample size, are usually subject to rigourous quality control checks, and represent a good source to select counterfactuals and contol post hoc.
- 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 andthen 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 METHOD DESIGNS

Interviews, focus groups and case studies are the most commonly used methods to collect qualitative data (Figure 16).

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 opinion. 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[35,36] and allow thorough profiling of a particular outcome. They can involve 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.

FIGURE 16 Qualitative Data Collection Tools

INTERVIEWS

- Interviews are conducted one one one, either face to face or over the phone and provide in depth understanding of the topic at hand as well as the opportunity to explore topics of interst in more depath and in real time.
- Standardised instruments through an interview schedule or protocol can be used as guides and questions may slightly differ to adapt to specific contexts (structured versus semi-structure interviews).
- Interviews can produce detailed data, but are resource intensive)both time and money).

FOCUS GROUPS

- Focus groups, like interviews are also conducted face to face using an interview protocol.
- They differ in that they are conducted in a small group of people with the aim to generate discussion, debate and for participants to express their opinions.
- Provide the opportunity for rich data to be collected for people to "bounce" ideaas off each other.
- Like interviews, however, focus groups can be relatively resource intensive.
- They are useful to test concepts (e.g. prior to a survey) or to explore concepts more in-depth (e.g. after a survey).

CASE STUDIES

- Used to illustrate good practice as well as provide contextual data.
- It involves an in-depth, possibly longer-term investigation of one or few participants in the program, and/or people interacting with them.
- Useful to understand a significant or novel situation.
- Risks include "cherry picking" – the researcher needs to ensure participants selected are likely to expose positive and negative outcomes.

MIXED OR MULTIPLE METHODS

Employing mixed or multiple methods for data collection (e.g. different types of quantitative and qualitative data collection techniques together) helps increase the accuracy of your measurement. Mixed methods can be used concurrently (e.g. open-end interviews conducted to affirm the validity of a survey) or sequentially (e.g. a focus group investigates topics that will be later explored in a survey, or a survey reveals matters that will be later explored through in-depth interviews/focus groups/case studies). 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!) the stakeholder analysis and your outcomes. It is important to understand who the information is about and who will you ask (e.g. you may ask the individual who achieved the outcome, but also their peers or family)
- ***What is the best instrument to collect the data?*** 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 or focus group may be more appropriate. See the table below.
- ***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, 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***[xv], 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, 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***.

Figure 21 in Appendix 3 presents a data collection matrix comprising various approaches (e.g. survey, interviews, self-assessment), the type of questions the method answers, how long the respective data collection exercise is likely to take, and aspects to consider when you decide to use a data collection approach. Complementing the outcomes and indicators developed in Section 5, Figure 17 presents a snapshot of the outcomes framework for **Sport**.

[xv] 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.)

FIGURE 17 Outcomes Framework – Outcomes, Indicators, Data Sources and Data Collection Time, Sport

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.

ACTIVITY #8-3: MEASURES AND DATA COLLECTION TIME

Complete the outcomes framework for your program.

Tips for outcomes framework development

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 scriber that leads the note-taking while the rest of the group brainstorms). Invite your team and key stakeholders if possible. 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. Move to the next column and add the indicator for this outcome (on a separate post-it note). 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 yourself organising the outcomes by short-, medium- and long-term or you might start by developing outcomes 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.

RESPONSIBILITY FOR DATA COLLECTION AND MONITORING

While data collection may be seen as essential to program activities and achievements by some stakeholders, it may be met with rejection by others who see it as consuming resources that could be otherwise allocated to 'doing good'.[21] Ensure you consider:

- **Who is responsible** for data collection, their understanding and capacity to collect the data – ensure staff have the skills and time allocated to collect the data.
- **Availability of participants** – your program participants are willing and available to provide you the information needed to quantify your outcomes.
- **Accuracy of data reported** – your tools are developed to capture the intended outcomes.[xvi]
- **Relevance of data collected** – 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.
- **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 confidentiality** – data should be kept on secure servers or locations that can be accessed only by the research team in a de-identifying manner, ensuring that individuals cannot be linked to their answers and their answers cannot influence their relationship with the program. The next section discusses ethical considerations in further detail.

ETHICS AND POLITICS OF DATA COLLECTION AND OUTCOMES MEASUREMENT

All research data collection requires ethical approval from a recognised committee. Different contexts have different formal requirements for ethics approval processes and it's important to know and understand these, to act upon ethical principles pertaining to human research and measurement. The NHMRC's (2018) National Statement on Ethical Conduct in Human Research[36] is the primary Australian framework to consider. Key principles include:

- **Integrity:** professionalism, excellence (using known, appropriate, and proportionate methods), honesty, reliability, stewardship.
- **Respect for persons and 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 disadvantage, being transparent about how participants are selected. Participation – or not – in measurement 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 measurement 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 any reporting.
- **Research merit and safety:** using sound and known methods, with quality assurance built in. Quality assurance might look like peer review, reference groups of experts, public communication.

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.

Remember that research takes place in a political context. It is important to also consider the ethical requirements for evaluators[39]:

- **Systematic enquiry:** assessment should be rigorous and include a discussion of limitations, not overclaiming.
- **Respect for people:** should respect the rights, privacy, confidentiality and dignity of all involved.
- **Competence:** adhere to research standards and rigour, reporting should be comprehensive and accessible.
- **Integrity/honesty:** disclosure of conflict of interest; report fairly and accurately.

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 the large-scale program and its purpose is to identify shortfalls and opportunities to improve the 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 little consent on the sample size necessary for a pilot study[37] 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[38]; a sample that is "representative of the population and sufficiently large, respectively".[37]

It is essential to evaluate the results of pilot studies, including outcomes and process 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 logic model.

The pilot should:

- be implemented according to the theory of change and logic model 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.

Good practice tip

It is good practice to collate the logic model and outcomes framework into a single LogFrame Matrix (see also LogFrame Matrix, SECO 2007).This will facilitate your understanding of the program from inputs to outcomes, the data collection needs and potential approaches to assess impact, as well as associated risks and assumptions. The LogFrame Matrix serves as a tool to plan, monitor and evaluate programs and projects. Turn to Appendix 4 for a LogFrame Matrix description and template.

[xvi] This refers to how the indicators are developed, and the data collection plan, including questionnaire design. CSI is developing a platform that will include an evidence base for key social issues. This will support organisations to collect rigorous data to measure social outcomes. See Amplify Social Impact: https://amplify.csi.edu.au/

7. ANALYSIS OF IMPACT

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 allocate to a single intervention, hence discussing ***contribution*** rather than ***attribution*** is often preferred.[43]

Steps for contribution analysis:[43]

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. direct beneficiaries) and externally – people knowledgeable about the program (e.g. local community members). There are some established techniques to isolate the impact of a program. They rely on measuring change compared to what would have happened had the program not been implemented. Turn to Appendix 5 to learn about techniques such as randomised controlled trials (RCTs) and quasi experimental designs (QEDs)[xvii].

DATA ANALYSIS

Data analysis will depend on the type of data 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 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 aspects of 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:

- Often qualitative data can be audio recorded; you need to transcribe the data (transfer the data in written format). There are specialized services who can do this for a cost.
- Code, analyse and write up the data:
  - » 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 mentioned by the participants themselves in the data, 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 research questions, evaluation terms of reference or outcomes framework etc).

[xvii] These are discussed below but see also the Magenta Book (HM 2011) for further details on RCTs and QEDs.

- Analysing the data means then 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, but is more commonly done using a computer software, such as NVivo.
- Braun and Clarke's (2006) 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, and (6) producing a write up of the findings.

Quantitative data analysis:

- Data collected in hard format (i.e. pen and paper) should be digitized (most often this means transferred in Excel).
- 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) has 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 Figure 24 in Appendix 5.

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 questions about the credibility of the evaluation and bias. Ensuring you have the right skills within your organisation is essential for rigorous and reliable measurement. This includes considering the skills and competencies of various people across an organisation, not only those that will be undertaking outcome measurement. Figure 18 presents five skill groups which are relevant for outcome measurement.

Reflection questions:

- What are the current skills and competencies in your organisation?
- What future skills and competencies will be required?
- What are the skill gaps?
- Are the skill requirements and implications understood?

FIGURE 18 Skills and Competencies for Outcomes Measurement

Skills and competencies	Description
Technical	These are predominantly hard skills required for: effective design of methods, implementation, data collection and analysis, interpretation and reporting. Thiscould include qualitative and quantitative methodologies.
Situational analysis	The skills required to understand, analyse and address the contextual and situational (political, economic, social and regulatory) issues around measurement.
Project management	The hard and soft skills required to manage a measurement project to completion. These include managing the measurement process, negotiating contracts, budgeting, identifying and coordinating resources, conducting the evaluation in a timely manner.
Interpersonal competencies	The interpersonal skills and emotional intelligence such as people skills, written and oral communication, negotiation, emotional intelligence and cross-cultural understanding.
Professional practices	The behaviour, norms and values that are foundational for evaluation practice, such as standards and ethics.

Source: Bennett et al 2016[31]

ACTIVITY #9: IDENTIFY THE TYPE OF ANALYSES AND PLAN FOR RESOURCES

8. COMMUNICATE IMPACT AND IMPLEMENT CHANGE

EFFECTIVE COMMUNICATION OF FINDINGS

How findings are communicated and used is as important as outcomes measurement itself. Effective communication will support accountability and learning[45] through communicating about results and communicating for results. Communicating about results is what is generally understood as communication of findings. It informs stakeholders about the findings of your evaluation. Communicating for results is also known as 'communication for development' or 'program communication' and is used as a management tool for internal learning and stakeholder engagement[45]. This type of communication focuses on internal learning, clarity across stakeholders and combined action.

The most effective communication techniques capture attention and interest, allowing audiences to interact with the findings.[46] Tailor findings to the audience and consider:

- Accuracy, balance and fairness
- Level of detail
- Technical writing style
- Appearance of the publication

Communicating negative or sensitive findings is an important aspect of communication and learning. Negative findings should be used for internal learning to redesign an intervention, improve approaches to interact with clients or deliver an activity. Results can point out groups of the target population for which an intervention worked as well as those for which it didn't, thus helping to identify 'pockets of disadvantage', groups or communities that are falling behind. This can help to develop tailored interventions to achieve better outcomes.

An implementation plan provides a summary of the process, roles, responsibilities and longer-term strategy to implement and administer your program's outcome measurement approach. An implementation plan has three key aspects[31]:

- **Integration:** establish outcome measurement processes within day-to-day activities and strategy. Alignment with existing frameworks, systems and tools.
- **Adjustment:** continual refinement and iteration of the outcome measurement approach, process, tools and methods.
- **Leadership and culture:** support a measurement culture for performance and continual learning.

? ***Reflection questions:***

- For what purpose are you conducting the analysis?
- Who needs to know, what and when?
- What level of detail?
- What format does the information need to be, e.g. data dashboards and visualisations, reports, or charts?

ACTIVITY #10: PREPARE YOUR COMMUNICATION AND IMPLEMENTATION PLANS

READY, SET, GO!

Evaluation 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, 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 is invaluable to understanding the impact of a program, the changes it makes to people's lives, how services can be improved, who is winning and who is missing out.

The Compass (Muir and Bennett, 2014) provided you with a checklist to understand if you are ready for outcomes measurement. We continue this checklist below with activities (which we discussed throughout the guide) that you should complete in your evaluation journey.

Writing an evaluation report

When writing an evaluation report, you must include at least the following sections:

- **Executive summary:** A high-level summary of the evaluation – what it did 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 is evaluated. The description of the project may be a separate section.
- **Evaluation framework:** Includes evaluation questions, scope, purpose, method. Describe the parameters of the evaluation – what questions you intend to answer, what is within the scope of the evaluation, the evaluation methods, and limitations.
- **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.
- **Conclusions and recommendations:** A high-level summary of the successes and lessons learned, as well as how findings should be used.
- **References:** The 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, questionnaires that were used for the data collection, more detailed results (for example further disaggregated by gender, or age groups).

Activities	What this means
1. ***Problem analysis and systems thinking***	Understand the complexities of the social issue your program is trying to resolve.
2. ***Clarify vision, purpose, mission, goals and objectives***	Understand what your program intends to achieve and how; explain how it fits within the organisation.
3. ***Stakeholder analysis: map your stakeholders and their engagement in the program and measurement***	Everybody counts. Map the groups engaged in your program, their role and responsibilities. Engagement of stakeholders for measurement.
4. ***Develop a theory of change***	Explain how your program will resolve the social problem.
5. ***Develop a logic model***	Further explain your theory of change by identifying the inputs, activities, outputs and outcomes of your program.
6. ***Formulate evaluation questions***	What questions do you want to answer?
7. ***Identify the evaluation type appropriate for your program***	Do you want to measure the outcomes your program achieves (outcomes evaluation)? Do you want to understand how activities can be improved (process evaluation)? When are you evaluating and why (formative evaluation, developmental evaluation, summative evaluation)?
8. ***Develop an outcomes framework***	Prioritise outcomes that you want to measure, develop indicators, identify data sources and timing of data collection.
9. ***Identify types of analyses and plan for resources. Collect and analyse data***	Plan the data analysis, including the resources to monitor, collect and analyse the data. Collect and analyse the data.
10. ***Prepare communication and implementation plans***	Write reports, brochures, pamphlets. Identify the best ways to communicate findings to a diverse audience. How will you use findings to improve your program, upskill or influence policy?

APPENDIX 1: OUTCOMES MEASUREMENT APPROACHES

Social Accounting and Audit (SAA) is a framework which enables organisations to build on existing documentation and reporting systems to account fully for and report on their social, environmental and economic performance and impact (prove); provide the information essential for planning future actions and improving performance (improve); and be accountable to all those they work with and work for (account).[47]

Social Accounting and Audit uses eight key principles:

- Clarify Purpose
- Define Scope
- Engage Stakeholders
- Determine Materiality
- Make Comparisons (benchmarking)
- Be Transparent
- Verify Accounts
- Embed the process

For further discussion of these principles and SAA, see www.socialauditnetwork.org.uk

Social Return on Investment (SROI) is a framework for measuring social, environmental and economic outcomes and uses monetary values to represent them. This enables a ratio of benefits to costs to be calculated. For example, a ratio of 3:1 indicates that an investment of $1 delivers $3 of social value.[48]

SROI was developed from social accounting and cost-benefit analysis and has seven principles:

- Involve stakeholders
- Understand what changes
- Value the things that matter
- Only include what is material
- Do not over-claim
- Be transparent
- Verify the result.

SROI can be evaluative (when completed at the end of a program) or forecast (completed prior to a program, initiative or policy being implemented). For a complete SROI guide see www.socialvalueuk.org

Social Return Accounting is a new framework to measure the value of projects ranging from physical infrastructure to social insurance schemes. The purpose of the approach is to develop a 'common language' for evaluating social returns from expenditure in a wide range of different areas so that spending on transport infrastructure, for example, can be more easily prioritised alongside something very different like education or healthcare. Social return accounting was developed in 2018 by UNSW Professor of Economics Richard Holden and UNSW Professor of Law Rosalind Dixon in partnership with economics consultant Alex Rosenberg. It expands the existing concept of social return on investment (SROI) by taking physical infrastructure into consideration as well as intangible forms of human or social capital. Full report is available here: http://www.grandchallenges.unsw.edu.au/sites/default/files/uploads/Social%20Return%20Accounting%20Final.pdf

Triple bottom line (TBL) proposes that companies should be preparing three bottom lines:

- The bottom line of the profit and loss account – the traditional measure of profit
- The bottom line of the "people account" – how socially responsible an organisation has been through its operations
- The bottom line of the "planet account" – how environmentally responsible the organisation has been.

The approach can be presented as the three Ps: profit, people and planet, looking to measure the financial, social and environmental performance of an organisation over a set period of time. The term and approach were coined in 1994 by John Elkington. In 2018 the author published an article titled '25 years ago I coined the term 'Triple bottom line'. Here is why it is time to rethink it'. See the links below for further information on the approach and the recent paper: https://hbr.org/2018/06/25-years-ago-i-coined-the-phrase-triple-bottom-line-heres-why-im-giving-up-on-it

Corporate Social Responsibility (CSR) has been coined in the 1950s and a range of definitions, depending on the fields of activity and research have been provided since. CSR is an organisational policy by which a firm or company complies with regulatory requirements and engages in actions that further social good, beyond the interests of the firm and that which is required by law. However, more recent advancement defines CSR as situation where the firm goes 'beyond compliance' and engages in actions that 'appear to further some social good beyond the interest of the firm and that which is required by law'.[49] The authors describe CSR activities to include:

- Incorporating social characteristics into products and manufacturing products
- Adopting progressive human resource management practices
- Achieving higher levels of environmental performance
- Advancing the goals of community organisations.

See McWilliams et al 2006[49] in the reference list for further discussions of CSR.

Results-based Accountability (RBA, also known as Outcomes-based accountability, OBA) is a framework that starts with the ends (what you want to create) and works back towards means (how you will create that change). It comprises six steps:

- What is the 'end'?
- How are we doing?
- What is the story behind the curve of the baseline?
- Who are the partners who have a role to play in turning the curve?
- What works to turn the curve?
- What do we propose to do to turn the curve?

For a complete guide on RBA, see https://clearimpact.com/results-based-accountability/

APPENDIX 2: USEFUL RESOURCES AND TOOLS

Below are provided a range of resources that you may wish to consult to deepen your skills in outcomes measurement.

Survey design

1. How to design a survey https://www.surveygizmo.com/resources/blog/designing-surveys/
2. Questionnaire design tips https://psr.iq.harvard.edu/files/psr/files/PSRQuestionnaireTipSheet_0.pdf
3. Principles of survey and questionnaire design http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Basic+Survey+Design+-+Questionnaire+Design

Interview guide design

4. Bryman, Allan 2012, Social Research Methods, Ch. 20 Interviewing in qualitative research
5. https://sociology.fas.harvard.edu/files/sociology/files/interview_strategies.pdf
6. https://msu.edu/user/mkennedy/digitaladvisor/Research/interviewing.htm

Quantitative data analysis

These tools present key aspects of data analysis, types of variables, as well as in-depth examples of within- and between-group analyses, and discussions of statistical significance, including tips for using statistical packages such as SPSS or Stata.

7. https://cirt.gcu.edu/research/developmentresources/research_ready/quantresearch/analyze_data
8. https://knowhow.ncvo.org.uk/how-to/how-to-analyse-quantitative-data-for-evaluation
9. https://cyfar.org/ilm_6_2 Data analysis, and quantitative data analysis (select from side-menu)
10. https://www.osii.nsw.gov.au/assets/office-of-social-impact-investment/files/Fact-Sheet-3-Randomised-and-non-randomised-designs-2018-July.pdf Randomised and non-randomised designs
11. Bryman 2012, Ch. 15 Quantitative data analysis
12. Bryman 2012, Ch. 16 Using IBM SPSS for Windows
13. Field (2009), Discovering Statistics using SPSS, Third Edition, Sage

Qualitative data analysis

These resources present qualitative data collection tools and analyses

14. https://research-methodology.net/research-methods/data-analysis/qualitative-data-analysis/
15. https://cirt.gcu.edu/research/developmentresources/research_ready/qualitative/analyzing_data
16. Bryman, Allan (2012). Social Research Methods, Ch. 17 The nature of qualitative research
17. Bryman, Allan (2012). Social Research Methods, Ch 24 Qualitative data analysis
18. Bryman, Allan (2012). Social Research Methods, Ch 25 Computer-assisted qualitative data analysis: using NVivo

Evaluation and outcomes measurement guides

Below are key evaluation handbooks and guides that offer key information on types of evaluation, principles of evaluation and tips on how to conduct an evaluation.

19. Kellogg Foundation Evaluation Handbook https://www.wkkf.org/resource-directory/resource/2010/w-k-kellogg-foundation-evaluation-handbook
20. The Magenta Book. Guidance for evaluation, HM Treasury 2011 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/220542/magenta_book_combined.pdf
21. The Green Book. Central Government guidance on appraisal and evaluation, HM Treasury 2018 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/685903/The_Green_Book.pdf
22. Office of Social Impact Investment (2018). Technical guide: outcomes measurement for social impact investment proposals to NSW Government https://www.osii.nsw.gov.au/assets/office-of-social-impact-investment/Technical-Guide-Outcome-measurement-2018-July.pdf
23. National Centre for HIV/AIDS, Viral Hepatitis, STD and TB Prevention. Types of evaluation. https://www.cdc.gov/std/program/pupestd/types%20of%20evaluation.pdf
24. University of Minnesota CYFAR. Types of evaluation. https://cyfar.org/different-types-evaluation
25. CSIRO (2020). Impact evaluation guide https://www.csiro.au/en/about/corporate-governance/ensuring-our-impact/evaluating-our-impact

26. 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 WA: Perth. DOI: https://doi.org/10.26182/gj21-wc02

Logic model development guide

27. Kellogg Foundation Logic Model Development Guide https://www.wkkf.org/resource-directory/resource/2010/w-k-kellogg-foundation-evaluation-handbook

Indicator banks

28. Social Progress Index, Amplify Social Impact https://amplify.csi.edu.au/social-progress-index/

29. Measures of Australia's Progress http://www.abs.gov.au/ausstats/abs@.nsf/mf/1370.0

30. Social Progress Index https://www.socialprogress.org/

31. OECD Better Life Index http://www.oecdbetterlifeindex.org/

32. OECD Society at a Glance (2016) https://read.oecd-ilibrary.org/social-issues-migration-health/society-at-a-glance-2016_9789264261488-en#page1

33. The World Bank https://data.worldbank.org/indicator

34. Charities evaluation services & National performance programme. Outcome and outcome indicator banks (2011) https://www.globalgiving.org/social-impact-academy-static/pdf/ces_outcomes_and_outcome_indicator_banks.pdf

35. Amplify Social Impact, Indicator Bank https://amplify.csi.edu.au/

Self-assessment tools

36. Readiness for Organisational Learning and Evaluation Instrument (ROLE) developed by FSG: https://www.fsg.org/tools-and-resources/readiness-organizational-learning-and-evaluation-instrument-role#download-area

37. Research in Practice for Adults. 2012. Organisational Audit for Evidence-Informed Practice. https://www.ripfa.org.uk/resources/publications/practice-tools-and-guides/organisational-audit-for-evidenceinformed-practice-2012/ For a small purchase price, this short tool comes with clear instructions for how to use across the organisation and across time.

38. Stewart, J. (2014). Developing a Culture of Evaluation and Research. Australian Institute of Family Studies. Child Family Community Australia (CFCA). The paper provides a number of indicators of what a culture of measurement looks like, including assessing leadership and other structures of readiness: https://aifs.gov.au/cfca/publications/developing-culture-evaluation-and-research/introduction

39. Muir, K. and Bennett, S. (2014). The Compass: Your Guide to Social Impact Measurement. Sydney, Australia: The Centre for Social Impact. See section 6 http://www.csi.edu.au/media/uploads/CSI_The_Compass.pdf

40. From Evaluation Support Scotland the 'Making it Stick Wheel' allows you to record progress and change in the process of embedding evaluation in your organisation: http://evaluationsupportscotland.org.uk/media/uploads/resources/making_it_stick_embedding_evaluation_wheel_with_explanation.pdf

41. Preskill, H. and Mack, K. (2013). Building a Strategic Learning and Evaluation System for Your Organization, FSG. The report suggests creating a vision for evaluation (measurement), and provide questions to consider to inform this process. See section 2: https://www.fsg.org/publications/building-strategic-learning-and-evaluation-system-your-organization

Stakeholder engagement tools

42. A short piece from McKinsey Consulting on how to take discussion of measuring impact to the Board: https://www.mckinsey.com/global-themes/leadership/the-four-questions-to-ask-when-serving-on-a-nonprofit-board

43. The Cancer Australia n.d. 'Consumer Involvement Toolkit' has multiple resources on consumer involvement, including service delivery and research. While focussed on issues to do with cancer, the knowledge is applied, accessible and transferable. There are checklists and templates to use for consumer engagement: https://consumerinvolvement.canceraustralia.gov.au/researchers/involve-consumers

44. Ontario Centre of Excellence for Child and Youth Mental Health. 2013. Program Evaluation Toolkit, http://www.excellenceforchildandyouth.ca/sites/default/files/resource/toolkit_program_evaluation_tools_for_planning_doing_and_using_evaluation.pdf

45. Preskill, H. & Jones, N. (2009). A practical guide for engaging stakeholders in developing evaluation questions. Robert Wood Johnson Foundation. https://www.rwjf.org/content/dam/farm/toolkits/toolkits/2009/rwjf48595

APPENDIX 3: TECHNICAL APPENDIX

FIGURE 19 Stakeholder Analysis – Level of Engagement

Level of engagement	Goal	Communication	Nature of relationship	Example
Passive	No goals, no engagement	No active communication	No relationship	Protests; concerns expressed through media
Monitoring	Monitor stakeholders' views	One-way communication, from stakeholder to organisation	No relationship	Media and internet tracking
Informing	Inform or educate stakeholders	One-way, from organisation to stakeholder, no obligation of stakeholder to reply	Short- or long-term relationships	"we will keep you informed"; letters; reports
Transacting	Work together in a contractual relationship (one party usually provides funding and directs the objectives)	Limited two-way: setting and monitoring performance according to terms of contract	Relationship terms set by contractual agreement	Funders; grant-making
Consulting	Gain information and feedback from stakeholders to inform decisions made internally	Limited two-way: organisation asks questions and the stakeholder answers	Short- or long-term involvement	Surveys; focus groups; one-to-one meetings; stakeholder advisory groups
Co-designing	Work directly with stakeholders to ensure their concerns are understood and considered in decision making	Two-way or multi-way between organisation and stakeholders	Short- or long-term engagement	Advisory panels, stakeholder forums, participatory decision making
Collaborating	Partner with or convene a network of stakeholders to develop mutually agreed solutions and joint plan of action	Two-way or multi-way between organisation and stakeholders. Stakeholders work together to take action. Learning, negotiating, decision making for all	Long-term	Joint projects, partnerships
Empowering	Delegate decision-making on a particular issue to stakeholders	Accountability is transferred to the stakeholders; stakeholders have formal role in governance and decision making	Long-term	Stakeholders are members of the governance structure (as members, shareholders, committees)

Source: Adapted from Krick et al. 2005[22]

FIGURE 20 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 is 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 clients would be to be asked this information?
Appropriate	How relevant do you think this indicator is for your clients' situation?
Feasibility	How practical would it be to collect this information from your clients?

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
Standardized test assessment	"How much have individuals changed, in their 'objectively' assessed 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"? "What might be causing this issue or facilitating these outcomes (because you have no preconceived ideas)"?	Medium – 30–90 mins	• What level of quality does your data need to have? How accurate does your data need to be? • Who are your desired respondents?	Depends on the accuracy of your data and whether it took a sample or census approach
Focus groups	"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"?	Medium – 30–90 mins	• 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 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

FIGURE 21 Data Collection Approaches Matrix

Administrative data (e.g., case management data)	"How many people are in the system"? "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

Source: Bennett et al. 2016[31]

FIGURE 22 Outcomes Framework Template and Example

Outcome	Indicator	Measure	Source and timeline for data collection
The change you intend to achieve in the short-, medium- and long-term	**The established indicator**	**How the indicator is being quantified; the survey question will you include in a questionnaire**	**Who you will be collecting this information from and when**
Improved mental health	Kessler 10	In the past 4 weeks, about how often did you... • Feel tired out for no good reason • Feel nervous • Feel so nervous that nothing could calm you down • Feel hopeless • Feel restless or fidgety • Feel so restless you could not sit still • Feel depressed • Feel that everything was an effort • Feel so sad that nothing could cheer you up • Feel worthless *(on a 1 to 5 scale, 1=none of the time; 2=a little of the time; 3=some of the time; 4=most of the time; 5=all of the time)*	Participant survey, pre-participation and 6-month during program

APPENDIX 4: LOGFRAME MATRIX

The LogFrame Matrix brings together components of the logic model (inputs, activities, outputs, outcomes, impact) and the outcomes framework (indicators, data sources). It is most often used in international development projects and is a useful tool to plan and monitor a project and measure its impact. Similar to the logic model, the LogFrame Matrix also includes risks and assumptions associated with a program or intervention, at all stages. This includes, for example, assumptions related to inputs (such as the ability to access some premises), but also risks associated with some activities (such as participants not being able to access some information due to some unforeseen restrictions).

It is important to include such risks and assumptions in the logic model or the LogFrame Matrix because they will help plan mitigation strategies, rework activities to achieve intended outputs and outcomes, and understand why some outputs and outcomes might not be achieved.

FIGURE 23 Logframe Matrix

	Component	Indicators / measures	Data sources	Risks / assumptions
Goal/Impact				
Outcomes	Long-term:			
	Medium-term:			
	Short-term:			
Outputs				
Activities		Inputs:		
				Preconditions

Adapted from SECO (2005).

The LogFrame Matrix has a horizontal logic and a vertical logic. The horizontal logic refers to the relationship across rows, for example between outcomes, indicators and their data sources. The indicator must be a good measure for the outcome and the data source must offer sufficient information to quantify the indicator. The vertical logic zigzags from the bottom-right corner of the LogFrame Matrix to the top-left corner, or from preconditions to impact. This is like the 'if-then' relationship of the logic model, discussed in Chapter 3. If the preconditions exist and the inputs are provided for the program, then activities can be conducted. If the activities are conducted, and the associated assumptions hold (and risk do not eventuate), the outputs will be obtained. If the outputs are achieved and the associated assumptions hold (and risks do not eventuate), the short-term outcomes will be achieved. This logic travels to the top-left corner, where impact will be achieved if the long-term outcomes are achieved and the associated assumptions hold and risks do not eventuate (or are well dealt with). For a further discussion of this logic, see SECO (2005).

APPENDIX 5 – RCT AND QED

Randomised Controlled Trials (RCTs)

RCT is a scientific experiment looking to minimise bias when assessing change due to a program. 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 and again during, at the end, and/or some time after the program completed. Analysing the differences between the two groups prior and post program will allow conclusions about the outcomes that can be attributed to the program. 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[xviii] 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)[xix].
- 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 two groups prior to the program, then again post-intervention. When designing a QED ensure that you are aware of potential threats to the integrity of the data:

- **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 didn't).
- **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).

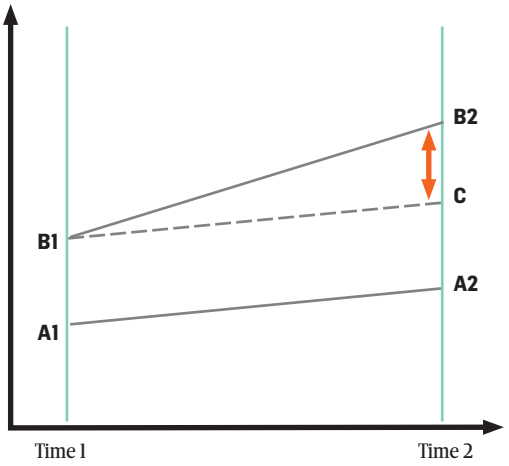
Source: Social Research Methods, n.d.[44]

[xviii] See Magenta Book, Paper 7, p.7.7 for further discussion of RCT and limitations

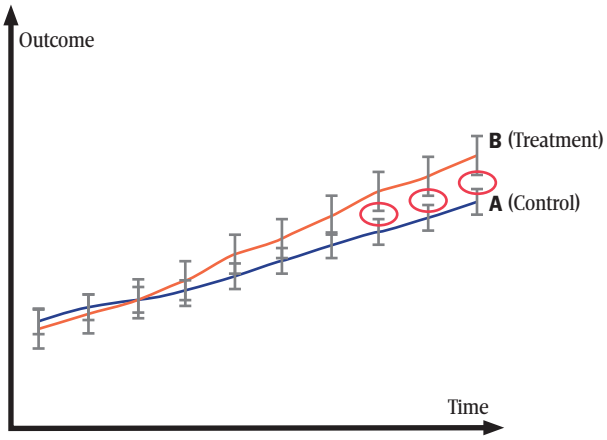
[xix] There are several techniques to overcome this issue, for example phased introduction (a program is rolled out in waves), intermittent applications (when interventions are very short-terms and may be repeated at different intervals across different groups), or accidental delays (such as implementing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

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

FIGURE 24 Example: Assessing Change (Quantitative Data)



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.

GLOSSARY

Activities: The processes or actions that produce the desired outputs and ultimately outcomes. In essence, activities describe 'what we do'.

Attribution: Attribution is the extent to changes can be considered a direct result of a program, initiative or intervention.

Baseline: The initial information collected about the condition or performance of subjects prior to the implementation of an intervention or program, 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.

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.

Evaluation: An objective process of understanding how a program, policy or other intervention was implemented, what effects it had, for whom, how and whyl. In an evaluation, social research procedures are systematically applied to assess the conceptualisation, design, implementation, and utility of programs or interventions.

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.

Logic model: 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.

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 measurement: A systematic way to assess the extent to which a program has achieved its intended results.[2]

Outcomes evaluation: The assessment of the changes resulting from the implementation of a program, policy or other intervention. It includes both intended and unintended outcomes for a range of stakeholders engaging in a program or intervention.

Outcomes framework: A collection of selected outcomes, indicators to measure those outcomes and the data sources necessary to quantify those indicators. It also includes the timing for data collection to measure change in each outcome.

Outputs: The direct products or services resulting from a program or intervention'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 richly description, 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).

Process evaluation: The investigation of the extent to which a program or intervention was implemented as planned. It helps understand why changes occurred.

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 impact assessment: The processes of analysing, monitoring and managing social impact.[3]

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.

Social impact: The intended and unintended social consequences, positive and negative, of programs (interventions, policies, plans, projects) and any social change processes invoked by these.[3]

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.

Systems thinking: Understanding a whole system – e.g. the social system – by examining the links and interactions between the components.

Theory of change: an explicit theory or model of how a program will achieve the intended or observed outcomes.[25] 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.

REFERENCES

1. 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.
2. 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.
3. Vanclay, F. (2003). International Principles For Social Impact Assessment. Impact Assess. Proj. Apprais. 21, 5–12.
4. Organisation for Economic Co-operation and Development – Development Assistance Committee (2010). Glossary of Key Terms in Evaluation and Results Based Management. Paris: OECD-DAC. http://www.oecd.org/development/peer-reviews/2754804.pdf.
5. Australian Institute of Family Studies (2018). Program planning and evaluation guide. Child Fam. Community Aust. https://aifs.gov.au/cfca/expert-panel-project/program-planning-evaluation-guide/export [Accessed September 6, 2019].
6. Muir, K., and Bennett, S. (2014). The Compass: Your Guide to Social Impact Measurement. Sydney, Australia: The Centre for Social Impact.
7. State Secretariat for Economic Affairs (2005). The Logical Framework User Manual. Bern: SECO. https://www.seco-cooperation.admin.ch/dam/secocoop/de/dokumente/resultate/monitoring/quali-manual-framework.pdf.download.pdf/Logical%20Framework%20-%20User%20Manual.pdf.
8. Bronfenbrenner, U. (1994). Ecological Models of Human Development. In International Encyclopedia of Education. Elsevier: Oxford.
9. Elias, A.A., and Cavana, R.Y. (2000). Stakeholder Analysis for Systems Thinking and Modelling. In. Wellington, New Zealand.
10. Wheelen, T.L., and Hunger, J.D. (2008). Strategic Management and Business Policy 11th ed. Upper Saddle River, NJ: Pearson Prentice Hall.
11. Hull, C.E., and Lio, B.H. (2006). Innovation in non-profit and for-profit organizations: Visionary, strategic, and financial considerations. J. Change Manag. 6, 53–65.
12. David, F.R. (2009). Strategic management: concepts and cases 12th ed. Upper Saddle River, N.J: Pearson Education.
13. Marjolin, A., Muir, K., and Powell, A. (2015). The Travel Companion: Your guide to working with others for social outcomes. Sydney, Australia: The Centre for Social Impact.
14. Mayne, J. (2009). Building an evaluative culture: The key to effective evaluation and results management. Can. J. Program Eval. 24, 1–30.
15. Conley-Tyler, M. (2005). A fundamental choice: Internal or external evaluation? Eval. J. Australas. 4, 3–11.
16. Wholey, J.S. (1979). Evaluation: Promise and performance. Washington, DC: The Urban Institute.
17. Wholey, J.S. (2011). Exploratory Evaluation. In Handbook of practical program evaluatio, J. S. Wholey, H. P. Hatry, and K. E. Newcomer, eds. San Francisco, CA: Jossey-Bass, pp. 81–99.
18. Stewart, J. (2014). Developing a culture of evaluation and research. Melbourne, Victoria: Child Family Community Australia, Australian Institute of Family Studies.
19. Inspiring Impact (2013). The Code of Good Impact Practice. https://www.inspiringimpact.org/wp-content/uploads/2018/01/Code-of-Good-Impact-Practice-1.pdf.
20. Ontario Centre of Excellence for Child and Youth Mental Health Program Evaluation Toolkit. Ottawa, Ontario: Ontario Centre of Excellence for Child and Youth Mental Health. http://www.excellenceforchildandyouth.ca/sites/default/files/resource/toolkit_program_evaluation_tools_for_planning_doing_and_using_evaluation.pdf [Accessed October 31, 2018].
21. Callis, Z., Flatau, P., and Seivwright, A. (2017). Progress in Outcomes Measurement: Findings from the Second Annual Outcomes Measurement in the Western Australian Community Sector Survey. Western Australia: Bankwest Foundation.
22. Krick, T., Forstater, M., Monaghan, P., and Sillanpää, M. (2005). From Words to Action: The Stakeholder Engagement Manual. AccountAbility, United Nations Environment Programme, Stakeholder Research Associates Canada Inc. http://www.unep.fr/shared/publications/pdf/WEBx0115xPA-SEhandbookEn.pdf.
23. Cancer Australia (2012). Consumer Involvement - working together for better outcomes: Tools for Researchers. https://consumerinvolvement.canceraustralia.gov.au/assets/involve/files/doc-lib/accessible/researchers/booklet_researchers_201212.pdf [Accessed October 31, 2018].
24. Patton, M.Q. (2012). Budgeting for Developmental Evaluation (DE). https://www.betterevaluation.org/sites/default/files/Budgeting%20for%20Developmental%20Evaluation.pdf.
25. Rogers, P.J. (2008). Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions. Evaluation 14, 29–48.
26. Connell, J.P., and Kubish, A.C. (1998). Applying a theory of change approach to the evaluation of comprehensive community initiatives: Progress, prospects, and problems. Washington, DC: The Aspen Institute.

27. Harries, E., Noble, J., and 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/.

28. Coryn, C.L.S., Noakes, L.A., Westine, C.D., and Schröter, D.C. (2011). A Systematic Review of Theory-Driven Evaluation Practice From 1990 to 2009. Am. J. Eval. 32, 199–226.

29. Norton, M.I., Mochon, D., and Ariely, D. (2012). The IKEA effect: When labor leads to love. J. Consum. Psychol. 22, 453–460.

30. W.K. Kellogg Foundation (2017). The Step-By-Step Guide to Evaluation: How to Become Savvy Evaluation Consumers. Battle Creek, Michigan: W.K. Kellogg Foundation. https://www.wkkf.org/resource-directory/resource/2017/11/wk-kellogg-foundation-step-by-step-guide-to-evaluation.

31. Bennett, S., Reeve, R., Muir, K., Marjolin, A., and Powell, A. (2016). Orienting your journey: An approach for indicator assessment and selection. Centre for Social Impact, UNSW, Sydney.

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

33. Drucker, P.F. (1986). The practice of management 1st Perennial Library ed. New York: Perennial Library.

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

35. Denzin, N.K., and Lincoln, Y.S. eds. (2013). The Landscape of Qualitative Research 4th ed. Los Angeles: SAGE Publications.

36. National Health and Medical Research Council (Australia), Australian Research Council, and Universities Australia (2018). National Statement on Ethical Conduct in Human Research 2007 (Updated 2018).

37. Johanson, G.A., and Brooks, G.P. (2010). Initial Scale Development: Sample Size for Pilot Studies. Educ. Psychol. Meas. 70, 394–400.

38. Hertzog, M.A. (2008). Considerations in determining sample size for pilot studies. Res. Nurs. Health 31, 180–191.

39. Australian Evaluation Society (2013). Guidelines for the Ethical Conduct of Evaluations. AES. https://www.aes.asn.au/images/stories/files/membership/AES_Guidelines_web_v2.pdf.

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

41. InspiringCommunities (2009). Developmental evaluation as alternative to formative assessment. https://youtu.be/Wg3IL-XjmuM.

42. Scriven, M. (1991). Evaluation thesaurus 4th edition. Newbury Park, Calif.: Sage Publications.

43. Mayne, J. (2001). Addressing Attribution Through Contribution Analysis: Using Performance Measures Sensibly. Can. J. Program Eval. 16, 1–24.

44. Social Research Methods (n.d.). Multiple Group Threats. http://www.socialresearchmethods.net/kb/intmult.php [Accessed September 15, 2019].

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

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

47. Kay, A. (2011). Prove Improve Account: The New Guide to Social Accounting and Audit. UK: Social Audit Network.

48. Nicholls, J., Lawlor, E., Neitzert, E., and Goodspeed, T. (2012). A Guide to Social Return on Investment. The SROI Network.

49. McWilliams, A., Siegel, D.S., and Wright, P.M. (2006). Corporate Social Responsibility: Strategic Implications*. J. Manag. Stud. 43, 1–18.

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

51. Waters Center for Systems Thinking (2020). https://ttsfilestore.blob.core.windows.net/ttsfiles/habits-single-page-2020.pdf#page=1&view=FitH

52. Dass, S., Muir, K. and Rangan, K. (2020). Nonprofit Boards: It Is Time to Lift Your Gaze and See the System. Working Paper 21-058. Harvard Business School.

CENTRE
for SOCIAL
IMPACT

UNSW
SYDNEY

www.csi.edu.au
Phone: *+61 2 8936 0909*
Email: *csi@unsw.edu.au*