



An Australian Government Initiative

Women in STEM Ambassador

EVALUATING STEM GENDER EQUITY PROGRAMS

A guide to effective program evaluation

Isabelle Kingsley



Author

Isabelle Kingsley

Office of the Women in STEM Ambassador, The University of New South Wales

Acknowledgements

This guide was produced by the Office of the Women in STEM Ambassador. The Office is an Australian Government Initiative funded by the Department of Industry, Science, Energy and Resources.

The author acknowledges the assistance of colleagues from the Department of Industry, Science, Energy and Resources in producing this guide.

December 2020

CONTENTS

Introduction	1
Why is evaluation of STEM gender equity programs important?	2
Evaluation basics	3
What is evaluation?	3
Four areas to evaluate	4
About this guide	4
How this guide is organised	5
Practical worksheets and templates	6
Your Turn	6
Icons used in this guide	6
See one, do one: A concrete example	6
Science in Australia Gender Equity (SAGE) Pilot	8
Define	11
Problem: what issues does the program address?	12
Audiences: who are the program and evaluation for?	13
Program audiences: who is the program for?	13
Evaluation audiences: who is the evaluation for?	14
Goals: what change do you want to create?	15
Outcomes: what short- to medium-term changes do you want to see?	15
Impacts: what long-term changes do you want to see?	16
Plan	18
Activities: what will you do to achieve the outcomes?	19
Inputs: what is needed?	20
Outputs: what is delivered?	21
Evaluation: how will you know what is achieved?	22
Evaluation priorities: what is most important?	22
Key questions: what do you want to know?	24
Indicators: what demonstrates that change is happening?	25

Design	27
Evaluation design: how will you carry out your evaluation?	28
Data collection: what data will you collect and how?	30
Choose a method	30
Instrument: what tool will you use to collect data?	35
Create your instrument	35
Run a pilot test	39
Data management plan: how will you safeguard data?	39
Execute	41
Recruit participants: invite people to take part in your evaluation	42
Collect the data: gather, document and record information	43
Analyse the data: examine, summarise and explain the data	44
Analysing quantitative data	44
Analysing qualitative data	44
Analysing mixed methods data	44
Share	49
Report the results: organise, summarise and present the findings	50
Transparency is key	50
Consider your audience	51
Organise your report	51
Disseminate your report: share your evaluation far and wide	53
Make it public	53
Share with key stakeholders	53
Consider publishing	54
References	55
Glossary	56
Appendices	57
Appendix A: Evaluation Planning Tool	58
Appendix B: Data management plan	59
Appendix C: Participant Consent Form Templates	60
Appendix D: Sample Focus-group protocol	62
Appendix E: Checklist	63



Introduction

Why is evaluation of STEM gender equity programs important?

"If you can't measure it, you can't improve it." — *Women in STEM Decadal Plan* (p. 23)

Underrepresentation of girls and women in STEM is a broad, complex and long-standing issue. Girls and women face many barriers to participating in STEM education and careers—from stereotypes and bias to inequitable workplace culture [1, 2]. STEM gender equity programs seek to dismantle these barriers to attract and keep girls and women in STEM. The end goal: a diverse and gender-balanced STEM workforce to confront important world issues [1, 3].

Australia has hundreds of programs to tackle STEM gender equity, but are they working? Unfortunately, there is little evidence that they do. Few programs are formally evaluated [3]. For example, a study conducted in 2018 identified at least 337 programs across Australia. Of those, only seven had publicly available evaluation results [4]. The lack of evaluation makes it difficult to know whether programs are effective and which programs should be extended or scaled-up [3].

Evaluation of STEM gender equity programs is a subject of national interest. This is why evaluation is a priority recommendation of the [Women in STEM Decadal Plan](#) [3]. Subsequently, the creation of this evaluation guide is a key action of the Australian Government's [Advancing Women in STEM 2020 Action Plan](#) [1].

Why is it important to evaluate STEM gender equity programs? Simply put: If we are going to try to create change, then we need to know if our actions are working.

SIX REASONS TO EVALUATE

Measure project outcomes. Evaluation is essential to gauge the effectiveness of your program. It helps you know whether your program is doing what it was meant to do.

Use resources efficiently. Evaluation allows you to justify and account for your program resources and investment. Knowing what works and does not work helps you use resources efficiently.

Attract and keep funding and partners. Evaluation results can help you attract and keep sponsors and partners. Results can show that the program is effective, valuable and worth investing in.

Improve and scale-up. Evaluation allows you to refine your program to improve it. For Government and sponsors, it informs decisions on what programs should be extended or scaled up.

Generate useful data. Evaluations can help us understand the issues that affect girls and women in STEM. Evaluation data over time helps us learn what programs work, for whom and why.

Promote best practice and collaboration. Evaluation can inform and guide best practice for STEM gender equity programs. Sharing evaluation findings can promote collaboration between programs.

Evaluation basics

This guide provides advice on how to evaluate your STEM gender equity program. Before getting started, it is important to cover some fundamental evaluation basics. This section provides a brief overview of evaluation. Then, it explains the four focus areas assessed in program evaluations.

What is evaluation?

Evaluation is collecting information to assess the strengths, weaknesses and effectiveness of a program [5]. The goal is to understand what works or does not work, how, for whom, and why.

Evaluation is not something that happens at the end of the program. It is an *integral part of the program* and is embedded within the program from the very beginning.

Four areas to evaluate

Evaluations should assess four main program areas: Design, Efficiency, Outcomes and impacts, and Lessons learned. This guide mainly focuses on evaluation of **outcomes and impacts**; but, the other three areas are also addressed throughout the guide. Try to address all four in your program evaluation.

Outcomes and impacts. Measure the change you want to achieve. Outcomes and impacts are the positive or negative changes produced by the program. The changes can be direct or indirect, intended or unintended [6, 7]. Assess both the extent and nature of the change [8]. Did the program do what it was meant to do? To what extent? Were there any unintended consequences? You can evaluate outcomes and impact throughout the program or at the very end.

Design. Think about what the program does, how it is done, who it is for and how it aligns with the goals of the program. How well does the program design address the targeted STEM gender equity issue? Is it targeting the right audience? Does the evaluation assess the outcomes and impacts you want to measure? You can evaluate the design throughout the program or retrospectively.

Efficiency. Monitor your program's resources and activities to catch any administrative problems or shortcomings. Did the program run smoothly? Did the program have enough, or the right, resources? Was it completed on time and budget? If not, why not? You can evaluate efficiency at any time or continuously throughout the program.

Lessons learned. Reflect on the experiences, strengths, challenges and outcomes of the program. What went well? What didn't go well? What can be improved, and how? Collect this information throughout the program and pull it together at the end.

About this guide

This guide offers practical advice to help you—STEM equity program leaders and evaluators—evaluate your program. It does not cover every nuanced aspect of evaluation. Instead, it gives in-depth guidance on targeted concepts. This guide focuses on evaluating program **outcomes and impacts**.

Program evaluation does not need to be daunting, difficult, expensive or time-consuming. This guide breaks it down into a simple, 5-step approach that's easy to apply to your program. It shows you how to embed evaluation within your program from the very beginning.

If you are picking up this guide at a later stage of your program and evaluation, do not fret. The guide is still useful. Use it from where you are in the process. Note that it is still worthwhile to do all the exercises for future planning and/or program improvements.

How this guide is organised

This guide contains five distinct sections based on a 5-step evaluation approach: Define, Plan, Design, Execute and Share (Figure 1). The colour-coded sections guide you through the five steps from beginning to end.

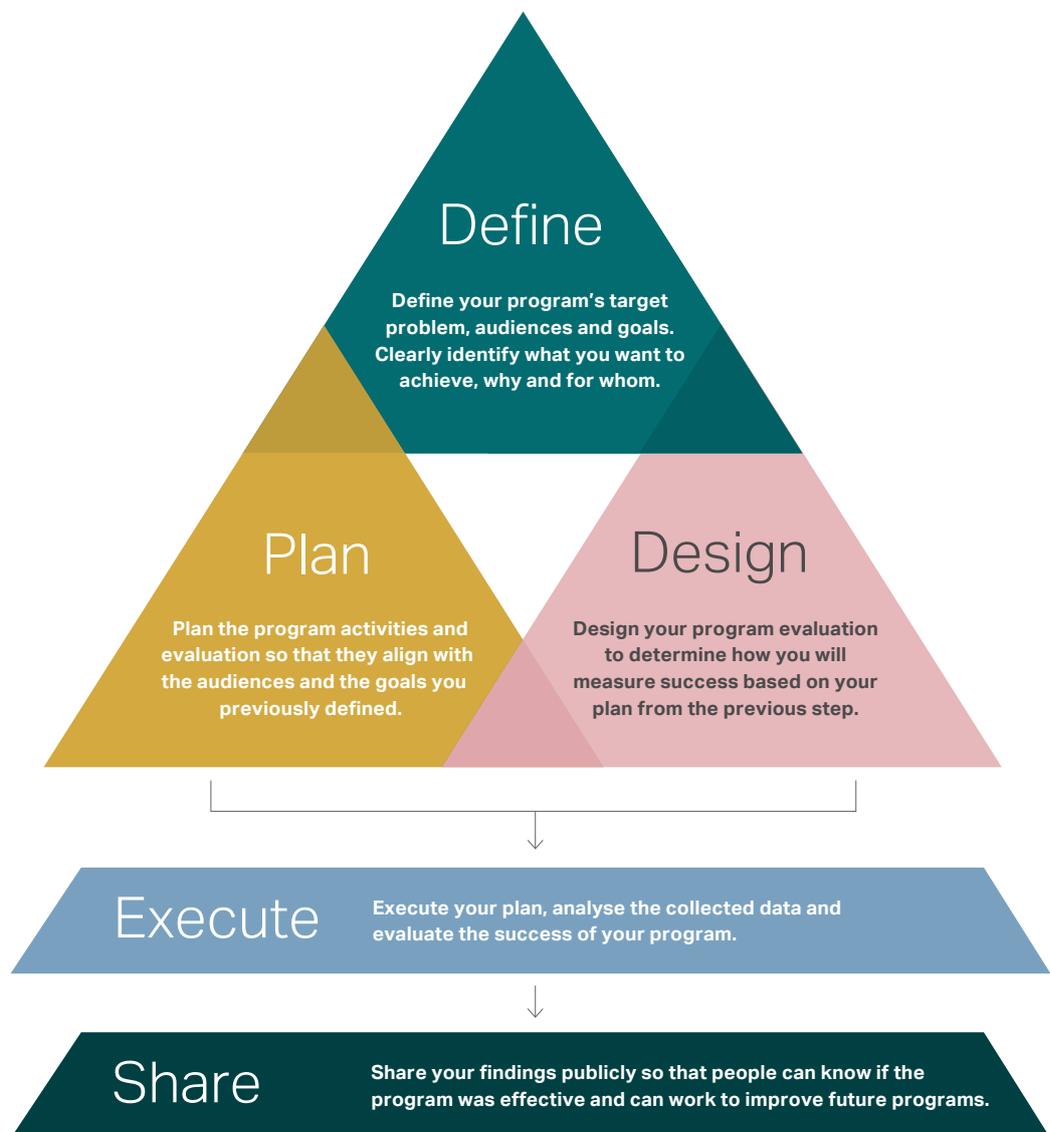


Figure 1. The five sections of the guide represent the five steps of the evaluation process.

Practical worksheets and templates

This guide has practical worksheets and templates in the Appendices section. They are useful tools to help you plan, design and conduct your program evaluation. Use them as-is or change them to suit your program. As you go through the guide, you will be prompted to complete the worksheets and templates (see 'Your Turn' below).

You can also use them to clearly and visually communicate your program and evaluation with colleagues and stakeholders in meetings, presentations and reports.

Your Turn

Your Turn boxes appear at the end of each section of this guide. They prompt you to stop and action the ideas in that section by completing a worksheet or template in the Appendix. They also summarise the key points and questions to consider.

Icons used in this guide

Two icons appear throughout this guide. The icons let you know that a specific tip or principle applies to the paragraph next to it.



Take note. This icon highlights noteworthy points, ideas and tips.



Remember to be sensitive and consultative. This icon highlights important sensitivity and consultation considerations for gender equity program evaluations. They are based on the UOW Framework for Evaluation of Equity Programs [9].

See one, do one: A concrete example

It's always helpful to have a concrete example to use as a model—see one, do one. Figure 2 is a completed *Evaluation Planning Tool* worksheet* for the Science in Australia Gender Equity (SAGE) Pilot program. The program evaluation is described in detail on the following pages.

*A blank version of the worksheet is available in Appendix A.

1. DEFINE

<p>PROBLEM What issues does the program address?</p> <hr/> <p><i>Poor representation of women in senior roles in STEMM HER institutions, due to structural and organisational barriers (e.g. lack of flexible work arrangements, sexual harassment, stereotypes, flawed recruitment and promotion practices, etc.)</i></p>			
<p>AUDIENCE</p> <hr/> <p>Program participants Who is the program for? <i>STEMM HER institutions applying to gain accreditation</i></p> <hr/> <p>Evaluation audience Who is interested in the eval.? <i>SAGE governing bodies and its management, STEMM sectors, government, academics, researchers, students</i></p>	<p>GOALS SMART / Specific - Measurable - Achievable - Relevant - Time specific</p>		
	<p>Outcomes - Short-term (immediate to 1 year)</p> <p><i>Pilot the UK Athena SWAN Charter in Australia with participating institutions (2015-2016)</i></p> <p><i>Raise awareness of gender equity and diversity issues in STEMM HER (2015-2016)</i></p>	<p>Outcomes - Medium-term (2-4 years)</p> <p><i>Implement learnings and recommendations from pilot (2017)</i></p> <p><i>Expand the pilot to make it available to all publicly-funded STEMM HERs in Australia (2017-2020)</i></p>	<p>Impacts - Long-term (5+ year)</p> <p><i>Reach the entire STEMM HER sector (2020+)</i></p> <p><i>Improve the participation, retention and success of women and other gender minorities and support gender diversity in STEMM HER in Australia (2020+)</i></p>

2. PLAN

<p>ACTIVITIES</p> <p>Activities What will participants do? <i>STEMM HERs adopt the 10 key principles of the Athena SWAN Charter and devise action plans that embed them within their policies, culture and practices</i></p> <p><i>Establish an internal SAT to analyse root causes of inequities within their institute</i></p> <p><i>Submit information for peer review (and accreditation by SAGE)</i></p>	<p>Inputs What is needed? <i>Guidance materials, resources, workshops, training programs, etc.</i></p> <p><i>SAGE team and management board</i></p> <p><i>Expert advisory group and SAGE peer review panels</i></p> <p>Outputs What is delivered? <i>SAGE subscribers</i></p> <p><i>Subscriber Accreditation Reports</i></p> <p><i>Athena SWAN Bronze Awards to institutes who successfully complete the process</i></p>	<p>Eval. PRIORITIES</p> <p>1-3 priorities for evaluation</p> <p><i>Evaluate the effectiveness of the pilot</i></p> <hr/> <p><i>Evaluate the Athena SWAN framework and accreditation processes to the Australian context</i></p> <hr/> <p><i>Evaluate the impact and benefits from the implementation of the framework in Australia</i></p>	<p>KEY QUESTIONS</p> <p>What questions will eval. answer?</p> <p><i>What went well? What didn't go well? What can be improved and how?</i></p> <hr/> <p><i>Is the framework appropriate for Australian HERs?</i></p> <hr/> <p><i>What are the impacts and benefits of SAGE for Australian HERs?</i></p>	<p>INDICATORS</p> <p>What demonstrates the outcomes?</p> <p><i>Notes, comments and feedback from organisations</i></p> <hr/> <p><i>The extent of HER buy-in and willingness to participate in the program</i></p> <hr/> <p><i>Increases in: career satisfaction, working practices, visibility of women and proportion of women in STEMM</i></p>
---	--	---	--	---

3. DESIGN

<p>DESIGN APPROACH</p> <p>Pretest, posttest, time series? <i>Posttest (by ACER)</i></p>	<p>METHOD</p> <p>Qualitative, quantitative, mixed? <i>Qualitative</i></p>	<p>DATA COLLECTION TOOL</p> <p>Survey, interview, observation? <i>Interviews, workshops and focus groups with senior leaders from SAGE member institutions, peak bodies, etc.</i></p>
--	--	--

Adapted from W.K. Kelloggs (2004) and Biggs and Tang (2011)

Figure 2. Completed Evaluation Planning Tool worksheet for the SAGE Pilot described on p.8–10. Worksheet adapted from W.K. Kelloggs [10] and Biggs & Tang [11].

Science in Australia Gender Equity (SAGE) Pilot: Gender equity accreditation process for the higher education and research sector

The SAGE Initiative piloted the Athena SWAN Institutional Bronze Award of the Athena SWAN Charter. Its aim was to “Improve gender equity in STEMM (STEM + Medicine) in the Australian higher education and research sector by building a sustainable and adaptable Athena SWAN model for Australia”. It did this by encouraging higher education and research (HER) institutions to adopt an accreditation process that identifies and addresses structural barriers and organisational culture. SAGE is an adaptation of the successful Athena SWAN Charter in the UK. The UK Athena SWAN Charter, through its ten principles, provides a framework for organisations to plan and enact work to advance gender equity under a recognition and award scheme, offering bronze, silver and gold level awards. The SAGE adaptation integrates gender diversity and inclusion into its accreditation process.

PROBLEM.

Women are poorly represented in senior roles in HER institutions. Studies show that women academics and researchers in STEMM are squeezed out of science careers by structural barriers and organisational culture. These barriers include: limited access to career development opportunities, lack of flexible work arrangements, sexual harassment, stereotypes and flawed recruitment, progression and recognition practices, among others.

AUDIENCES.

The **program audiences** are STEMM HER institutions who apply to gain SAGE accreditation for Athena SWAN Institutional Bronze Award by undertaking a structured self-assessment and reflective examination of institutional policies, practices and data guided by the 10 key principles of the Charter.

The **evaluation audiences**—those interested in knowing or applying the evaluation findings—include many stakeholders: SAGE governing bodies and its management, government, Australian STEMM sectors as well as academics, researchers and students.

GOALS.

In the first instance (in 2015-2016), SAGE will pilot the adapted UK Athena SWAN Charter in Australia with participating institutions and raise awareness of gender inequity in the STEM HER sector (**short-term outcome**). From 2017, the learning and recommendations from the pilot will be implemented and the pilot will be expanded to all publicly-funded STEM HER institutions in Australia (**medium-term outcomes**). From 2020, the SAGE's goal is to reach the entire STEM HER sector to drive systemic, structural and cultural organisational changes. The aim is to improve the participation, retention and success of women and other gender minorities and support gender diversity in STEM HER in Australia (**long-term impact**).

ACTIVITIES.

STEM HER organisations who participate in SAGE must adopt the 10 key principles of the Athena SWAN Charter, undertake a structured evidence-based approach and devise action plans that embed these principles within their policies, practices, and culture to achieve Bronze accreditation. In doing so, they must establish an internal cross-disciplinary self-assessment team (SAT) to perform an analysis of the gender equity and diversity to understand the root causes of inequities within their institute and devise an actionable four-year plan. SATs then submit this information to be assessed by a broad range of expert peer reviewers, in a process managed by SAGE and overseen by independent moderators. SAGE then accredits institutions via the Athena SWAN Award to successful institutions.

The **inputs** for these activities include: guidance materials and resources, specialist workshops and training program delivered by SAGE and regional network meetings enabled by funding from the Australian Government and subscription fees paid by participating institutions, and an annual symposium, SAGE team and management board, an expert advisory group and the SAGE peer review panels who assess applications.

The **outputs** from the activities include: SAGE subscribers (HER institutions who are applying for accreditation) applications for accreditation and Athena SWAN Bronze awards to institutions who successfully complete the accreditation process.

EVALUATION.

The Australian Council for Educational Research (ACER) was engaged to undertake the evaluation.

The three **evaluation priorities** include:

1. Evaluate the appropriateness and effectiveness of the pilot implementation undertaken to from 2015 to 2017 (Design; Efficacy)
2. Evaluate the Athena SWAN framework and accreditation processes to the Australian context for use in Australian tertiary education and research institutional environments (Design).
3. Develop and establish a data framework to inform future assessment and reporting on impact and benefits realisation from the implementation of the Athena SWAN framework in the Australian context (Outcomes and Impacts; Lessons Learnt).

The first priority sought to answer the following key evaluation questions: What went well? What didn't go well? What can to be improved and how? Key questions and indicators were identified for each of the evaluation priorities (see *Evaluation Planning Tool* on pg. 7 above).

Between April and October 2017, ACER conducted widespread consultations involving more than 140 people, including senior leaders from SAGE member institutions, peak bodies, government agencies and a range of external stakeholders. This involved individual interviews, regional workshops and on-site focus groups (**posttest, qualitative**).

A precis of ACER's report '[Evaluating the introduction of Athena SWAN into Australia: putting gender on your agenda](#)', communicates the key findings of ACER's evaluation of the SAGE Pilot. It was publicly published on the SAGE website and shared with stakeholders, funders and government.

The report outlines ACER's examination of the value of SAGE and Athena SWAN in Australia's STEMM HER sector, details next steps and recommendations for future actions.



Define

Define your program's target problem, audiences and goals. Clearly identify what you want to achieve, why and for whom.

The define step sets the stage for the program and its evaluation. It prompts you to ask yourself:

- What is the problem that your program addresses?
- Who are the program and its evaluation for?
- What do you want to achieve?

This section helps you define your program's target problem, audiences and goals. Articulating these key elements will allow you to have a clear idea of what you want to achieve, why and for whom.

If you are picking up this guide midway or at the end of your program, you may be tempted to skip ahead. Resist the urge. No matter where you are in the process, clearly defining these elements will focus your evaluation and reporting.

Problem: what issues does the program address?

Identify what problem your program wants to address. There are many different factors, challenges and systemic barriers that contribute to the underrepresentation of girls and women in STEM.

Examples include:

- Stereotypes, such as gendered views of career choice and caring roles
- Lack of flexible work arrangements to enable caring and other responsibilities
- Bias, both conscious and unconscious
- Workplace culture, such as harassment, exclusion, and progression practices

Specifying what problem(s) your program addresses gives your program a clear purpose. It justifies **why** your program is needed. Having a clear purpose, as a first step, helps ensure you build your program to tackle the problem to achieve the change you want to create (outcomes and impacts). It will help you define your program audiences and goals and will focus your activities on what is likely to elicit the change you want.





Take a look at [Appendix 1 of the Women in STEM Decadal Plan](#) [3].

It outlines common barriers that impact girls' and women's participation in STEM in Australia. Your program likely aims to target one (or a few) of these common barriers. If your program specifically targets problems related to young girls, take a look at the [Barriers to Participation in Engineering and the Value of Interventions to Improve Diversity](#) report [2].

Audiences: who are the program and evaluation for?



Program audiences: who is the program for?

Describe the intended audience of your *program*. Who is the program for? Consider the problems you want to address and identify who is involved.

Program audiences are the people participating in your program. They can include women working in STEM fields, young girls in school, teachers, parents and carers, STEM organisations or policymakers.

Example:

Take the [STEM Women](#) online directory as an example. The program addresses two main problems:

1. Problem 1: The underrepresentation of women role models, mentors and voices at STEM events, on committees and in the media.
2. Problem 2: Lack of access to career development and progression opportunities.

In this case, there are two main audiences involved in the program:

1. Women in STEM: Women working in STEM who want opportunities to be mentors, speakers or committee members to raise their profile and advance their careers.
2. Seekers: Organisations or individuals seeking STEM expertise or looking to meet gender equity targets or increase the visibility of women in STEM.

Note. Take a look at the full example of the STEM Women program evaluation at [womeninstem.org](#).

Know your audiences. Is there demographic data about your program audiences or similar ones? What can you find out about their needs, interests, values, attitudes or knowledge? STEM gender equity reports and policy documents have information about audiences of various equity issues and gaps.

Evaluation audiences: who is the evaluation for?

You also need to define the audiences of your *evaluation*. Who is interested in the evaluation findings?

Evaluation audiences can include program staff and internal departments in your organisation. They may also include external stakeholders such as program partners, sponsors and funders. The program participants, their families, current and future staff and the local community may also be interested. Government agencies are interested in your findings (see the 'Government: A key audience' box below).

Example:

Take the example of the STEM Women online directory (above). The program is run by the Australian Academy of Science and is funded by the Australian Government. In this case, the evaluation audiences include a variety of stakeholders. The evaluation findings are of interest to the Academy of Science and the Government. They are also of interest to its users, program partners and organising committee, as well as policymakers and future partners or funders.

Defining who the evaluation is for will help you determine its priorities (see p. 22). It will also help you tailor your report and how you disseminate the findings (see p. 51).



Make sure that the program and its evaluation are culturally, socially and ethically sensitive to your audiences.

This means that the program is respectful to participants and stakeholders of different cultures, groups and backgrounds. Be aware of the values, beliefs, needs and behaviours of various social and cultural groups. Consult with participants and stakeholders. Make sure that there is agreement and mutual understanding. Use the [AIATSIS guidelines](#) to make sure your program and evaluation respect its participants and stakeholders.

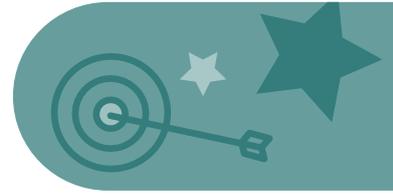
GOVERNMENT: A KEY AUDIENCE

Government is a key 'evaluation audience.' The Australian Government strategically targets investment and effort based on evidence of what works, and what is needed for the Australian context. Evaluation findings can help inform decisions about what programs should be funded or scaled up.

This is why this evaluation guide is a key action of the Australian Government's [Advancing Women in STEM 2020 Action Plan](#). The government has taken steps to make evaluation a condition of grant funding, including under the [Women in STEM and Entrepreneurship \(WISE\)](#) grants program. The government recommends the use of this guide to evaluate the success of current and future WISE projects in meeting their intended outcomes.

By understanding what works, government can strategically target future actions to address gender inequity.

Goals: what change do you want to create?



Start with the end in mind. What do you want to achieve?
What does success look like?

Focus on goals that create systemic and cultural change. Think beyond the ‘Band-Aid’ solution. Aim to treat the underlying cause of the problem that you want to address. Systemic and cultural changes are permanent changes to the way organisations, institutions or systems function to remove barriers to women and girls’ participation in STEM education or careers. Unless STEM gender equity programs lead to long-lasting or structural shifts in attitudes, behaviours, practices or policies, their effects will be temporary. Focus your goals on enduring, system-wide changes to tackle and break down the barriers.

Define your goals as specific and targeted **outcomes and impacts**. Outcomes and impacts are the changes you want the program to produce. Although often used interchangeably, there are important differences between outcomes and impacts.

Outcomes: what short- to medium-term changes do you want to see?

Outcomes are the short- and medium-term changes produced by the program.

- **Short-term.** Immediate changes (<1 year). They are the easiest to evaluate because you can assess them during and/or soon after the program.
- **Medium-term.** Changes that happen 2 to 4 years after the program. They may be more challenging to evaluate in short-lived programs. However, programs that run for multiple years, or that can track participants over time, can assess medium-term outcomes.

GET SMART

Goals tend to be abstract and lofty, such as “inspire girls to study STEM”. Measuring such goals is challenging because they are vague.

Define SMART outcomes and impacts:

Specific: Use specific language to describe what you want to achieve.

Measurable: Define what will change and by how much.

Achievable: Be realistic with what you can achieve with the time and resources available.

Relevant: Relate the program outcomes and impact to the problem or gap it addresses.

Time-specific: State the timeframe for the changes to happen.

Outcomes often refer to changes in knowledge, skill, interest, attitude and behaviour. Table 1 below outlines five common categories of program outcomes.

Impacts: what long-term changes do you want to see?

Impacts are the long-term changes that happen as a result of the program outcomes (5+ years). Impacts are broader in scope than outcomes. They are social or cultural changes in organisations, the community or the environment. Impacts are likely to be those that you have a less direct influence on. They are also much harder to evaluate (see the 'Knowns and unknowns' box below). One way to address this challenge is to define impacts as aspirational and focus your evaluation on short-term outcomes.

Three examples of the difference between outcomes and impacts:

- A program may want higher education and research institutions to be equitable workplaces in 10 years (impact) by adopting gender equity accreditation processes and principles now (outcome).
- A program may want to increase the diversity of STEM role models in 5 years (impact) by getting organisation leaders to pledge to meaningfully involve diverse women on panels now (outcome).
- A program may want to create a more diverse STEM workforce in 5 years (impact) by helping employers recruit qualified STEM professionals returning to work after a career break now (outcome).

Note. Take a look at the full examples of these program evaluations at womeninstem.org.

KNOWNNS AND UNKNOWNNS

There is a wide variety of STEM gender equity programs. These programs have many different intended *outcomes* that all contribute to shared future *impacts* on gender equity in STEM.

The issue is that many of the impacts will take decades to manifest. For example, you may want to know whether a program for young girls influences their career pathways. But, the impacts of that program won't be known for 10 years. During this time, a lot of factors can come into play. The girls will have many different experiences. The world and the nature of the workforce will change. These factors affect how much you can link impacts to a program. As such, the impacts of STEM gender equity programs are elusive and remain largely unknown.

How can you address this issue? Define the impacts as aspirational and focus your evaluation on the short-term outcomes. They are directly measurable or 'knowable'.

Remember: the link between the knowns (outcomes) and unknowns (impacts) is not easy to establish. Think about how to evaluate and report on outcomes and impacts to avoid the risk of seeming anecdotal.



Consult your potential audiences or relevant stakeholders when defining the outcomes and impacts. Consultation is a continuous process, but it's most important in the planning phase. It will help you make sure that the outcomes and impacts are appropriate and sensitive to the groups who may be affected by them.

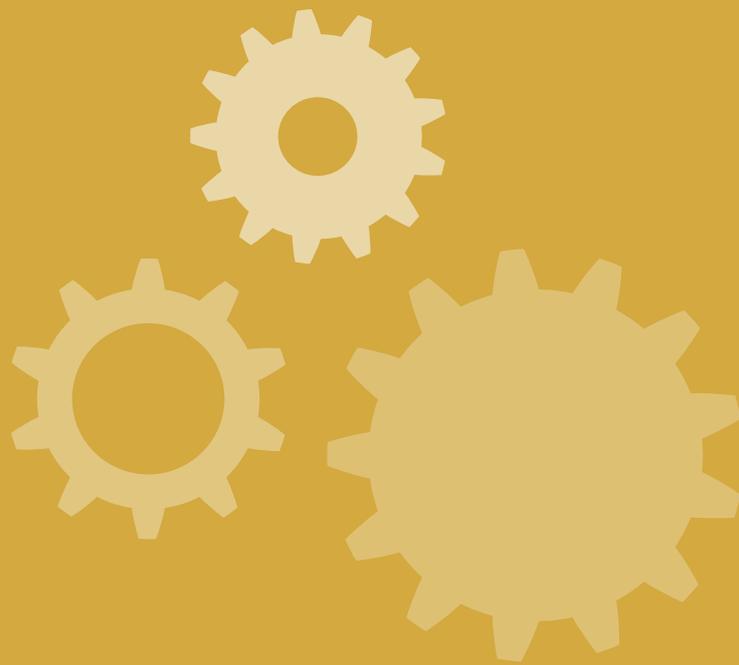
YOUR TURN

Define the problem, audiences and goals of your program. Use the *Evaluation Planning Tool* worksheet in Appendix A. Ask yourself: What problem(s) does your program address? Who is the program for? Who is interested in the evaluation findings? What do you want to achieve in the short, medium and long term? Make sure your outcomes and impacts are SMART. Use the sample evaluations at womeninstem.org.au to help you define your goals.

Table 1. Outcome categories and definitions.

OUTCOME CATEGORY	DEFINITION
Enjoyment & Inspiration	<ul style="list-style-type: none"> • Having fun • Innovative thoughts • Exploration, experimentation and making • Being inspired
Attitudes and values	<ul style="list-style-type: none"> • Feelings • Bias or stereotypes • Perceptions • Opinions about self (e.g. self-esteem) or others • Tolerance or acceptance • Openness to change • Empathy • Motivation • Positive and negative attitudes about an experience
Knowledge & understanding	<ul style="list-style-type: none"> • Knowing about something • Learning facts or information • Making sense of something • Deepening understanding • Making links and relationships between things
Skills	<ul style="list-style-type: none"> • Knowing how to do something • Being able to do new things • Intellectual skills • Information management skills • Social skills • Communication skills • Physical skills
Behaviour & progression	<ul style="list-style-type: none"> • What people do • What people intend to do • What people have done • A change in reported or observed actions • A change in the way people live or work • A change in a system, protocol or procedure
Other	<ul style="list-style-type: none"> • Program-specific

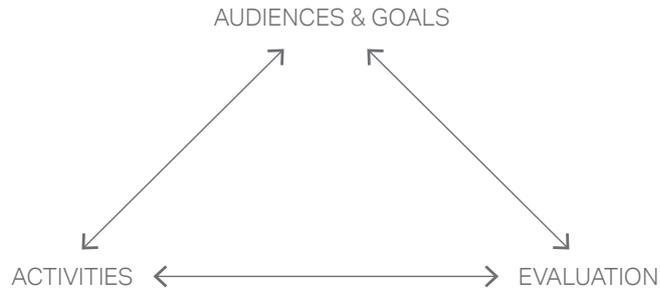
Adapted from the Generic Learning Outcomes, Arts Council England [12].



Plan

Plan the program activities and evaluation so that they align with the audiences and goals you previously defined.

The planning step is the most important part of the program and evaluation process. It allows you to embed evaluation into the program from the very beginning. In this step, you will choose the program activities and plan the evaluation.



The activities and evaluation must align with the goals and audiences so that they perform properly together¹. Think about the goal that you defined in the previous section and ask yourself:

- What will you do to achieve it?
- How will you measure what is achieved?

This section helps you plan the program activities and evaluation so that they align with the audiences and goals.

Activities: what will you do to achieve the outcomes?



Choose activities that are likely to elicit the outcomes (and eventual impact) you want to achieve.

An *activity* is a thing that a person or group does to achieve a specific outcome. There are a lot of different STEM gender equity programs with a wide range of intended outcomes. Their audiences are also diverse. As such, different programs can call for very different activities.



Depending on where you are in your program development stage, it may be appropriate for you to skip this step for now. Come back to it when making improvements to activities based on your evaluation findings, or when planning a whole new program.

¹ The planning step applies the principle of 'constructive alignment'. Constructive alignment is an approach used by educators to design courses and assessments—but, it also applies to programs and evaluations. At the simplest level, constructive alignment gets you to plan and design the different elements of your program so that they 'align' and perform properly together. It means that your intended outcomes for your audiences shape the design of the activities (what people do) and the evaluation (demonstration) of those outcomes [11].

Align your activities to the program audiences and goals (outcomes and impacts) that you defined in the previous section. Choose program activities that are likely to achieve the goals for your audiences.

Example:

Take the [STEM Returners](#) program as an example. The program **goal** is to return people to the STEM workforce after a career break. More specifically, it aims to change recruiters' negative views on CV gaps.

The **activities** align with these goals. The STEM Returners program consists of a 12-week work placement for STEM professionals who want to return to work after a career break. The paid placements are an alternative to direct hire recruitment. They allow companies to assess candidates on practical competencies and skills, rather than their employment history. Companies can monitor participants' progress and choose at the end of the placement, whether to hire a candidate permanently. Services and opportunities are available to the employers and employees to support them during the placement.

EVIDENCE-BASED ACTIVITIES

Base your activity choices on evidence—not on what feels intuitively good to do.

Seek out information to guide your activity choices. You may find a program that achieved the same or similar outcomes you want from your program. If so, what activities did they use? Is there research on the outcomes that you want to address?

Take a look at [Appendices 2-4 of the Women in STEM Decadal Plan](#) [3]. They outline research insights and suggested solutions (mapped to barriers) that can help inform your program activities.

Note. Take a look at the full example of the STEM Returners program evaluation at womeninstem.org

As you choose your program activities, consider the inputs and outputs. Inputs are the resources needed, and outputs are the direct products of your program.

Inputs: what is needed?

Inputs are the resources that go into your program. There are four main types of resources:

- time (how long things will take)
- human (staff, volunteers)
- financial (budget)
- material (consumables, infrastructure) resources

Identify all your inputs—they are part of the investment in achieving your goals. They are also part of evaluating your program's efficiency (described in the Introduction, p. 4).

Knowing your inputs will help you answer questions such as: Is the program adequately resourced to undertake its planned activities? Is the program on time and budget? Are the program outcomes and outputs worth the investment?

Outputs: what is delivered?

Your program activities will generate outputs. Outputs are the direct services, events, products and deliverables of your program.

Example:

Take the example of the [STEM Returners](#) program (above). Outputs from the program include services for employers and returning STEM professionals, such as:

- Mentoring and career coaching program
- CV writing and interview preparation workshops
- Networking events

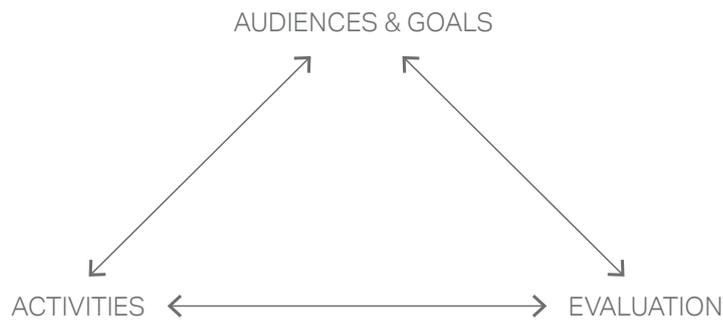
Identifying and articulating your intended program outputs is important to the evaluation. Outputs become your quantitative measures to monitor and report in your evaluation plan. Outputs are evidence that the program's activities were performed as planned.

YOUR TURN

Outline your program activities, inputs and outputs. Use the *Evaluation Planning Tool* worksheet in Appendix A. What will you do to achieve the outcomes? What resources are required? What products and deliverables will your program generate?

Evaluation: how will you know what is achieved?

In the previous section, you designed appropriate program activities that will elicit the outcomes you want. Now, you will identify how to measure if the activities achieve those outcomes.



Evaluation priorities: what is most important?

Identify the priorities of the evaluation. You can't evaluate every aspect of the program. Trying to assess too many things can result in evaluations that are incomplete or lack depth. Instead, identify one to three priorities for the evaluation. What do you want to know the most? What will provide the most useful information?

Align your evaluation priorities to the audiences, goals and activities.

Consider the following:

- Think about what **outcomes (and impacts)** you want to achieve. Then consider which ones you can realistically measure to assess your program. Long-term impacts may not be possible to measure for years. What is directly and immediately measurable? Refer to 'Knowns and unknowns' on p. 16.
- Think about the **purpose** of the evaluation and how you want to use the findings. To refine or further develop the program activities? To assess the program outcomes? Or, to justify a rationale for continued funding? It may be any or all of these purposes.
- Think about the **evaluation audiences** you identified in the previous section (p. 14). Who is interested in the evaluation findings? What information is of most value to those audiences?



Not all evaluation priorities will relate to your program outcomes and impacts. Some may relate to other areas of the program, such as the program design, efficiency and lessons learned (see the four areas to evaluate on p. 4).

Example:

Take the [Techgirls Competition](#) program as an example. It is a 12-week hands-on program for 7 to 17-year-old girls. Student teams are connected with women mentors and coaches to design solutions to real-world problems in the form of apps. They produce business plans, pitches and working app prototypes.

The intended **outcomes** for the students are to: improve their coding knowledge, confidence in using technology and increase their interest in STEM as a future career choice. The program also aims to provide an enriching experience for the mentors and coaches.

The purpose of the evaluation was to find out if the program was successful in achieving these outcomes. It was also to find out what worked and what didn't to refine and improve the program. The evaluation was intended for the program-owners, current funders and potential future funders (**evaluation audiences**).

Based on the outcomes, purpose and evaluation audiences, the **evaluation priorities** were to:

1. Measure changes in students' coding knowledge, confidence using technology and interest in a future STEM career.
2. Determine the motivations of the mentors and coaches to participate in the program and what they gained from participating.
3. Determine what worked, what didn't and how it could be improved.

Note. Take a look at the full example of the Techgirls Competition program evaluation at womeninstem.org.

YOUR TURN

List up to three priorities for your program evaluation. Use the *Evaluation Planning Tool* worksheet in Appendix A. Refer to the sample programs at womeninstem.org. Ask yourself these questions: What is most important? What outcomes are measurable? Who is interested in your findings, and what do they want to know? How will the findings be used?

Key questions: what do you want to know?

Once you have established your evaluation priorities, develop key evaluation questions around them.

The 'key questions' are the high-level questions that you want the evaluation to answer. They are not specific questions that you will ask program participants. The two are linked—the questions for the participants will help answer the higher-level key evaluation questions.

Here's an example of the difference:

A high-level key question that you want the evaluation to answer is:

- To what extent did the program increase girls' intentions to pursue STEM careers?

A specific question that you ask the program participants to answer the key question is:

- Pretest: I am considering a career in Science or Technology (strongly agree – strongly disagree)
- Posttest: Having participated in the competition, I am now more likely to consider a career in Science or Technology (strongly agree – strongly disagree)

This section focuses on the high-level key questions. You will develop the participants' questions later in the 'Design' section (p. 35).

Align your key questions to your evaluation priorities. Depending on your priorities, some questions may relate to outcomes and impacts and others to design, efficiency or lessons learned. Design and efficiency-related key questions can be simple 'yes' or 'no' questions. Outcomes-focused questions relate to the changes produced by the program. For these, use open-ended questions to allow you to assess the extent of the change.

Examples:

Take the example of the Techgirls Competition program mentioned above.

An example of an outcomes-focused question might be:

- To what extent did the program increase girls' confidence in using technology?

A design-focused question for the same program could be:

- Were the program activities appropriate for the target audience (girls 7 to 17-years-old)?

An efficiency-focused question may ask:

- Did the program run as planned?

Finally, a question related to lessons learned might ask:

- What worked well, what can be improved and how?

YOUR TURN

Write the key evaluation questions in the *Evaluation Planning Tool* in Appendix A. What do you want to know? Write specific and targeted questions. Use the [Better Evaluation](#) website—it has lists of key evaluation questions for the different types of evaluations.

Indicators: what demonstrates that change is happening?



Determine the indicators for each of the priorities and related key questions. 'Indicators' are markers that demonstrate whether, and to what extent, change is happening. They are usually an increase or decrease in a change your program wants to achieve.

Indicators are specific, observable and measurable. They can be:

Numbers. The number of women returning to work after a career interruption; the number of organisations awarded gender equity accreditation.

Percentages. The percentage of women enrolled in VET and university STEM subjects; the percentage of gender discrimination reported in an organisation.

Rates. The rate of men accessing parental and carer's leave; the rate of women entering the STEM workforce.

Proportions. The proportion of women and men in senior and/or leadership roles in an organisation; the proportion of research funding allocated to women applicants.

Specifying the indicators for your key questions will establish what you want to measure. It will also frame how you carry out the evaluation and collect data.



Check the [STEM Gender Equity Monitor](#) for ideas of indicators that could demonstrate change. It is an interactive national data report that shows the current state of STEM gender equity in Australia. It also provides a baseline for measuring change and trends over time, which could be useful as a comparison for your evaluation.

YOUR TURN

Outline the indicators that will answer the key questions you wrote in the previous step (use Appendix A). What will demonstrate the outcomes that you want to achieve and measure? Refer to the examples of program evaluations at womeninstem.org to help you.

The design stage determines how you will carry out the evaluation. You will base your design on the priorities, key questions and indicators that you determined in the previous section. First, you will decide on the type of data you will collect, when and how you will collect it, and how you will store it. Then, you will develop and test the tool that will collect the data.

This section is not a comprehensive module on research design and methodologies. The [Better Evaluation website](#) has detailed information about many different designs and methods. Instead, this section provides guidance on common, practical approaches for quality program evaluation.



Evaluation design: how will you carry out your evaluation?

There are many different ways that you can carry out your evaluation and collect data. The four common designs are: posttest, pretest/posttest, time series and retrospective pretest/posttest [13]. Here, the word 'test' refers to the instrument you use to collect the data. A 'test' can be a survey, task, interview or focus-group (discussed later in this section on p. 30).

Posttest. Data are collected after the program. Participants take part in the program and are tested afterwards.

Pretest/posttest. Participants are tested before (pretest) and after the program (posttest). The results of the pretest and posttest are compared. The difference between the two measures how much change the program achieved.

Time series. This is a variation of the pretest/posttest design. It collects data over regular intervals of time. Participants are tested before, during, and after the program. The results of all the tests are compared to measure change over time intervals.

Retrospective pretest/posttest. Another variation of the pretest/posttest design, but participants are tested only after the program. Participants report their initial ideas retrospectively (pretest) and their current ideas (posttest). For example, participants may be asked how they felt about their skills before the program and how they feel about them now.

Table 2. Strengths and weaknesses of common evaluation designs

DESIGN APPROACHES	STRENGTHS	WEAKNESSES
Posttest <i>program → posttest</i>	Simple to administer Provides data quickly	Does not measure change Does not measure whether the outcome is due to the program or to some other cause
Pretest/posttest <i>pretest → program → posttest</i>	Measures change over time	Cannot account for other factors that might have influenced or caused the outcome
Time series <i>program (test 1, 2...) → posttest</i>	Measures change over regular time intervals	Cannot account for other factors that might have influenced or caused the outcome Best used in longer programs
Retrospective pretest/posttest <i>program → pretest & posttest</i>	Useful when pretest is impossible Measures change over time using a one-time test Reduces burden on participants	Determining causal relationships may be difficult Respondents' self-reporting of change may bias the results Respondents may find it difficult to remember prior ideas or behaviours

Things to consider:

- The **strengths and weaknesses** of each evaluation design (Table 2).
- The **feasibility and suitability** of the design. Think about the context of your program, your resources and what you are comfortable with. Not all designs are practical and appropriate for all programs.
- Your **priorities, key questions and indicators**. If you want feedback about the program, consider a posttest to collect data from program participants, staff or stakeholders. Posttest evaluations are useful to assess program design, efficiency and lessons learned. If you want to measure changes produced by the program, consider a pretest/posttest or time series design. If a pretest is impossible, then use a retrospective pretest/posttest. Pretest/posttest design (and its variations) are useful to assess outcomes and impacts.



It is difficult to measure cause and effect—meaning, whether the program (cause) produced the change (effect). STEM gender equity programs are delivered in complex and unstable environments. A lot of other factors come into play that might influence or cause the change. You have limited control over them. This makes it difficult to link the program to the changes that you observe. To address this, measure the changes and give evidence of how they link to the program. When you report the results, account for the other factors that might have influenced or caused the change.

BEYOND THIS GUIDE

Experimental design is beyond the scope of this guide. Experimental design randomly assigns participants to a group. The groups either get a treatment (experimental group) or not (control group). Then the results of the groups are compared.

This design is rigorous and effective. But it moves beyond evaluation and into research. It is difficult to administer and requires a certain level of research expertise. It also requires compliance with particular processes (e.g. approval from a human ethics committee).

Some large-scale programs or those led by research organisations may choose to do this kind of research. But most programs are unlikely to be able to use experimental design. This guide gives advice on common and practical designs for quality evaluation.



Data collection: what data will you collect and how?

Think about **what** data you want to collect based on your key questions and indicators. Then, think about **how** you will collect the data. Consider the program activities, audiences, setting, atmosphere and timeframe.

Choose a method

There are three main methods to collect data: Quantitative, Qualitative and Mixed Methods. Each method involves different data collection tools, with different approaches, strengths and weaknesses (see Table 3). Choose the method that best suits your key questions and indicators.

Table 3. Strengths and weaknesses of common methods, tools and approaches

METHOD	TYPICAL DATA COLLECTION TOOLS	DESIGN APPROACHES	STRENGTHS	WEAKNESSES
<p>Quantitative</p> <p>Collection of quantitative or numerical data</p> <p>Analysed using statistics</p> <p>Results displayed as tables, charts, histograms, and graphs</p> <p>Findings tend to be broad and generalisable</p>	<p>Surveys or tasks that use close-ended questions</p>	<p>Choice questions</p> <p>Yes/no, multiple-choice</p>	<p>Simple and quick to administer</p> <p>Provides standardised responses across all respondents</p>	<p>Options are forced and may not reflect the true choices of respondents</p>
		<p>Rating questions</p> <p>Rating scales, Likert scales (agree/disagree), star rating, slider</p>	<p>Quickly measures preferences and opinions for individual questions</p>	<p>Does not determine why the rating was selected</p>
		<p>Ranking questions</p> <p>Ranking scales, drag-and-drop</p>	<p>Quickly measures order of preferences for a group of questions</p>	<p>Does not determine the reason for the preferences/order</p> <p>Take more time to answer</p>
<p>Qualitative</p> <p>Collection of data in the form of text, images, video and audio</p> <p>Analysis explores themes and trends</p> <p>Findings are richly detailed but less generalisable</p>	<p>Open-ended survey questions, observations, interviews, focus groups, work samples, journaling, mind mapping</p>	<p>Open-ended questions</p> <p>Respondents give an answer in their own words</p>	<p>Allows participants to share their ideas, thoughts and knowledge.</p> <p>Provides information about context (why/how)</p>	<p>Can be more complex to analyse</p>
		<p>Narrative</p> <p>Respondents tell their story using spoken, written or visual forms</p>	<p>Allows participants to share their own story</p> <p>Provides information about context (why/how)</p>	<p>Storytelling approach may have limited applicability</p>
		<p>Case study</p> <p>Examine a single person, group or event</p>	<p>Uses a variety of data collection methods to provide detailed accounts</p>	<p>Limited generalisability</p>
<p>Mixed methods</p> <p>Quantitative and qualitative methods mixed in one overall evaluation</p>	<p>Any mix of quantitative and qualitative data collection tools</p>	<p>Any mix of quantitative and qualitative design approaches</p>	<p>Combines the strengths of qualitative and quantitative designs</p> <p>Allows for triangulation</p>	<p>Can be more complex because it requires the analysis of two types of data</p>

Adapted from Phillips et al. 2014 [5].

QUANTITATIVE METHODS

Quantitative methods use numerical data that can be analysed using statistics. Quantitative data come from closed-ended questions—meaning, questions that have answer options for people to choose from. They include questions with rating scales, ranking, true/false, multiple-choice and demographic categories. To analyse them, you convert the choices and categories into numbers (e.g. true = 1 and false = 0). Then, you compute the numbers using statistics. The results are expressed as tables, charts, graphs and histograms. Quantitative methods are easy and inexpensive to administer, and the data is quick to analyse. They are useful when you need to draw broad or general conclusions from your evaluation. But, they do not provide detail about context.

There are three common approaches to collect quantitative data: choice questions, rating scales and ranking questions.

Choice questions. Allows respondents to select one or more options from a list of specific answers. Examples of choice questions include 'yes/no' or multiple-choice. You can display your choices using checkboxes or dropdown menus (for longer lists).

Rating scales. Displays a scale of answer options of any range (e.g., 0-100 or 1-10). Respondents select the number that best represents their response. You can also create rating scales using stars (star rating) or a slider. Likert scales are also a type of rating scale. Likert scales ask respondents how much they agree or disagree with something. The scale usually goes from 'strongly agree' to 'strongly disagree.'

Ranking questions. Asks respondents to rank their answer choices in order of preferences. Ranking allows you to assess how respondents feel about the options relative to each other.



Quantitative data can also be counts. Counts can be the number of women returning to work after a career break, the number of people at an event or the number of women and men promoted in an organisation.

QUALITATIVE METHODS

Qualitative methods use data in the form of text, images, audio and video recordings. Data can come from surveys, tasks, observations, journaling, mind maps and work samples.

There are three common approaches to collect qualitative data: open-ended questions, narrative and case studies.

Open-ended questions. Ask participants to give an answer in their own words. These questions don't provide specific pre-defined answer options. Open-ended questions are used in surveys, interviews and focus groups. They are also used in narrative and case study approaches (see below).

Narrative. Allow participants to tell their story using spoken, written or visual forms. Participants recount events or actions in their lives. It is used to collect data about social context and complicated events. It is often used in gender and education studies [14].

THINK OUTSIDE THE 'SURVEY' BOX

Surveys are often the first thing that comes to mind when we think of evaluation. Surveys are easy, and cheap ways to quickly collect data. But surveys have their drawbacks. They are overused, which leads to survey fatigue—people avoiding surveys because there are too many surveys. Survey results can also be affected by low response rates, biases, language barriers and question interpretation issues.

Think outside the box—be creative. What other sources of information can provide evidence of your program outcomes? Interviews, focus groups, and observations are common methods. However, think beyond conventional methods. For example, you can:

- Collect social media engagement content generated by program participants (e.g. tweets, posts, responses)
- Use a live polling app to collect participant responses to questions during your program
- Collect work samples or professional portfolios that participants produce during/for the program
- Ask participants to create concept maps (or mind maps) about x topic to visualise what they know
- Use an implicit associate test² (IAT) to measure biases about x topic (e.g. unconscious gender bias)

As much as possible, embed data collection into the program by making it part of the program activities. Consider the burden on your participants.



As much as possible, embed data collection into the program by making it **part of the program** activities. Consider the burden on your participants.

² Implicit association tests (IAT) detect the strength of a person's implicit, or subconscious, biases. They test a person's association between concepts (e.g. women, men) and judgments (e.g. pretty, smart) or stereotypes (e.g. artist, engineer). IATs get people to quickly sort words into categories and measures the response time. The faster the response time, the more ingrained and implicit the association. There are IATs for different topics available online (e.g. [Harvard University Project Implicit](#)).

Case study. A detailed, up-close and in-depth look at a single 'case'—a person, group, event or community. It is useful to explore issues, decisions, policies and systems. Case studies use many different approaches together (e.g. observations and interviews).

Qualitative methods provide detail about the context of what you are measuring—*what, how, when* and *how much*. For example, qualitative methods can help you understand *why* an outcome was or wasn't achieved. Or they can help you understand *how* a program can be improved.

Qualitative methods let you delve deep into your key questions. When you analyse qualitative data, you look beyond the numbers and explore the ideas, themes and trends in the content. Qualitative methods let you explore complex and nuanced outcomes like attitudes, values and motivations.

MIXED METHODS

Quantitative and qualitative methods work better together. Using them together is called a mixed methods approach. In mixed methods, you collect data from different sources and 'triangulate' your findings. 'Triangulation' means that you compare the findings from various sources to see if they support the same conclusions. Triangulation gives strength and confidence to your evaluation findings. It also helps build a more complete picture of what you are evaluating.

Example:

A program evaluation seeks to assess if a workshop increased teachers' understanding of STEM careers. It uses three methods: (1) a closed-ended pretest/posttest survey question (quantitative); (2) an open-ended pretest/posttest survey question (qualitative); and (3) tweets about the workshop (qualitative). The findings:

- The close-ended question shows a 12% increase in teacher's understanding.
- The open-ended question reveals that teachers' initial ideas about STEM careers (e.g. lab coats and robots) broadened to include different disciplines (e.g. trades and agriculture).
- The teachers' tweets contain lots of comments about STEM careers being different in the future.

The three different data sources support the same conclusion: the workshop increased teachers' understanding of STEM careers. Together, they paint a more holistic picture of the outcomes of the program. The quantitative data provide concrete numerical measures. The qualitative data provide contextual detail about how teachers' understanding increased and what they learned. Notably, the qualitative data uncovered insights that may have been missed using only quantitative data.



Make sure that the evaluation design, methods and approaches you use are culturally, socially and ethically sensitive. This means that they are respectful to participants and stakeholders of different cultures, groups and backgrounds. Consult with potential participants and stakeholders to get their advice and input.

YOUR TURN

Outline the evaluation design(s), method(s) and data collection tool(s) you will use in your evaluation in the *Evaluation Planning Tool* in Appendix A. What do you want to know? What methods will provide the most useful data to answer your questions? Consider your indicators (p. 25). What markers demonstrate whether, and to what extent, change is happening? Use Tables 2 and 3 to help you decide.

Instrument: what tool will you use to collect data?



In the previous section, you decided what data collection method and tool you will use. Now, you will **develop** it. This step involves creating an instrument to collect the data. It also involves testing it to make sure it works as planned and measures what it is intended to measure.

Create your instrument

Data collection tools are often referred to as 'instruments.' Instruments can be protocols for interviews, focus groups or observations. They can be surveys, live polls or any task that you want the participants to complete.

You have three choices. You can either use an existing instrument, modify an existing instrument or develop your own.



Although you have three options, it is highly recommended that you **use an existing instrument**. If no instrument is suitable, then consider modifying one. Creating your own instrument is discouraged, and it should be a last resort.

USE AN EXISTING INSTRUMENT

Use an existing instrument that has been tested and shown by its developers to be valid and reliable. Valid and reliable means that the instrument accurately and consistently measures what it is meant to measure. Validity and reliability bring rigour to your evaluation and credibility to your findings.

There are a plethora of instruments that measure the outcome categories (Table 1, p. 17) relevant to STEM gender equity programs. There are instruments to measure attitudes towards STEM, interest in STEM careers and STEM identity. There are instruments to measure perceptions of STEM and gender bias. For example, a quick internet search for 'interest in STEM instrument' returns over 100-million hits in a fraction of a second. Many instruments are also specific to particular audiences (e.g. 'schoolteachers' or 'recruiters').

How do you decide amongst the sea of choices?

Ask the following five questions:

1. Does the instrument address your **key questions**? Does it measure your **indicators**?
2. **Who developed** the instrument? Do they have the appropriate expertise?
3. Has the instrument been tested and shown to be **valid and reliable**?
4. Is the instrument appropriate for the **audience**? (e.g. age, education level, culture)
5. How is the instrument **administered**, and how is the data **analysed and interpreted**?

Selecting an existing instrument requires some research and decision-making. However, it is easier, faster and more rigorous than creating your own. It lets you compare your findings with other studies and evaluations that also used it. It makes your evaluation findings more credible.

MODIFY AN EXISTING INSTRUMENT

If you cannot find an existing instrument that suits your needs, then consider modifying one. This involves making changes by adding, removing, and/or tweaking items to make it more suitable. But, these changes might affect its validity and reliability. For this reason, this option is not recommended.

If you modify an instrument, make it clear in your evaluation report what changes you made to it and why. If you can, get in touch with the instrument developers. Check with them to make sure that your changes don't affect its validity and reliability.

Example:

A program leader wants to measure changes in perceptions about gender bias in STEM workplaces and attitudes towards equitable recruitment practices. They find an instrument to measure perceptions of gender bias. They find another to measure attitudes towards equity in staff selection. Each instrument contains over 20 questions. Together, the two instruments have too many questions. The program leader decides to modify them. They select the 5 most relevant questions from each instrument. They combine them to create a single instrument containing 10 questions.



Take a look at the example evaluation of the Techgirls Competition program at womeninstem.org. For a section of the survey, the program-owners used an existing instrument to measure interest in STEM careers called the *STEM Career Interest Survey (STEM-CIS)* developed by researchers in the field of STEM education.

DEVELOP YOUR OWN INSTRUMENT

Creating your own instrument is not recommended. It takes a lot of time, effort and expertise to create a quality instrument. Developing an instrument involves research, consultation, refining, revising and correcting. It also involves rigorous testing to make sure it is valid and reliable.

If you must create your own instrument, there are five important guidelines to follow:

1. **Think back to your questions and indicators.** What do you want to know? What markers demonstrate the change you want to achieve? Refer to the 'Think in terms of *what* and *how well*' box below.
2. **Do some research.** Find out how others measure similar outcomes. What approaches and instruments do they use?
3. **Consider your participants.** Who will use the instrument? Children? Adults? Think about education level, language barriers and cultural sensitivities.
4. **Produce a draft.** Write the questions, create the game or formulate the task. Produce whatever it is that you will use to collect data.
5. **Consult.** Share the draft with colleagues and experts in the field. Ask for feedback about the outcomes you are measuring, as well as language, clarity and length.

If the instrument you are developing is a survey or questionnaire, follow the [Australian Bureau of Statistics \(ABS\)](http://www.abs.gov.au) guidelines.

DEMOGRAPHIC QUESTIONS

Demographic information is important and useful to collect. Include demographic questions that will help you describe your participants. Common demographic questions include age, gender, geographic location, occupation, ethnicity, and education level.

A good idea is to align your demographic questions to existing standards, such as the [Australian Bureau of Statistics \(ABS\)](#). This makes it easier to compare your sample to the broader population. For example, by using the same standard [cultural/ethnic groups](#) as ABS, you can compare your participants to the Australian population.



Think about the participants in your sample. Consider any cultural sensitivities around collecting demographic data (e.g. indigenous population – refer to the [AIATSIS guidelines](#)).



Consult with stakeholders and potential participants to identify any sensitivities around the collection of demographics (and other) data.

THINK IN TERMS OF *WHAT* AND *HOW WELL*

You may want to measure changes in participants, such as their awareness, interest, understanding or skills. To capture these changes, think about what participants need **do** to demonstrate the outcome. Also, consider *how well* they need to do it. Some things will show a higher level of performance than others.

Use specific verbs based on the level of performance that you want to measure.



Example 1: An awareness campaign to improve carers' understanding of STEM careers

What do participants (carers) have to **do** to demonstrate 'understanding'? What level of understanding?

- **List** x number of STEM careers
- **Explain** what you think the term 'STEM' stands for

Example 2: A workshop to improve science communication skills of women scientists

What do participants need to do to show improved 'communication skills'? One measure may be to ask participants to assess their own ability.

- **Rate** your ability to communicate science to public audiences

Other measures may be to ask participants to demonstrate their ability by how well they can do a task.

- **Explain** your research to a 5-year old using no jargon
- **Describe** x phenomenon using a metaphor
- **Create** a 1-minute video about your most recent research journal article

Run a pilot test

Run a pilot test. A 'pilot' is a small-scale test of your design and instrument to make sure it works properly. Try to find people who are similar to those who will be using it in the evaluation.

If you put an existing instrument into an online survey or mobile polling app, have a few people run through it. Ask them to check that there are no technical glitches. If the instrument is a task (e.g. implicit association test or a mind map), have a few people perform it. Ask them if the instructions are clear. Observe to make sure that people actually do what you intended them to do. If you developed a protocol for interviews or focus groups, then do a mock run-through with a few people to test the questions, structure and flow.

If you developed your own survey, you must test it. You need to test whether it is valid and reliable—if it accurately and consistently measures what you want it to measure. Follow the [Australian Bureau of Statistics \(ABS\)](#) guidelines for testing surveys.

Data management plan: how will you safeguard data?



You must have a robust data management plan. Data management plans outline how you will collect, store and safeguard your data. This is of utmost importance to protect the privacy of your participants.

Consult the [Australian Privacy Act](#). It contains 13 [Australian Privacy Principles](#) around the collection, use, access and disclosure of personal information. Apply these principles to your evaluation.

Consider these questions: Where will digital and hard copies of data be stored? Who will have access to the data? How long will you keep it? If data are confidential and anonymous, what protocols are in place to make sure confidentiality and anonymity are maintained?

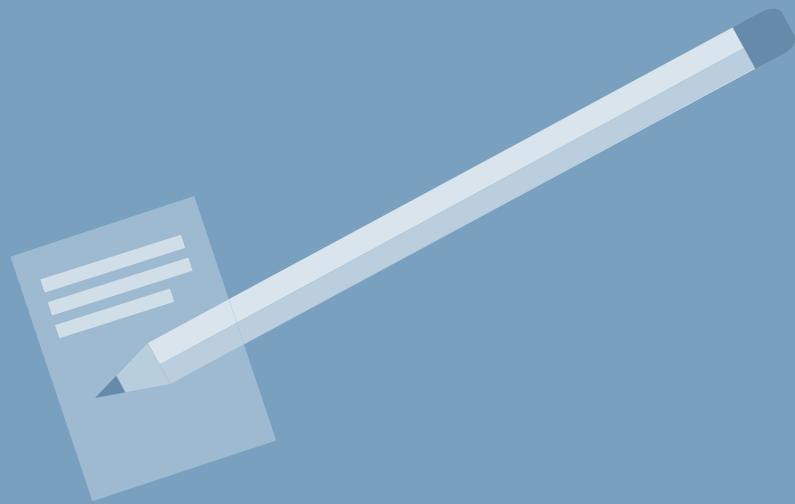
Table 4. Data management plan (with examples)

DATA TYPE	HOW IS DATA STORED SECURELY	HOW IS ACCESS RESTRICTED	HOW LONG WILL DATA BE KEPT
Hard Copies	e.g. Data is stored in a locked cabinet. Office/ Storage location is also locked with key access restricted to program staff	e.g. Only the program staff can unlock the cabinet/ office/storage room door	e.g. 5 years after reporting of findings OR 7 years after completion of the program
Electronic Copies	e.g. Data is encrypted in the cloud; a secure link is required to access and share the data	e.g. Only program staff will have access to the data	e.g. 5 years after reporting of findings OR 7 years after completion of the program
Audio/Visual	e.g. Data is encrypted in the cloud; a secure link is required to access and share the data	e.g. Only program staff will have access to the data	e.g. 5 years after reporting of findings OR 7 years after completion of the program

YOUR TURN

Create your instrument. Use your completed *Evaluation Planning Tool* (Appendix A) to guide you. What existing instruments are available for you to use? If creating your own, use the guidelines and ABS resources provided above.

Develop your data management plan using Table 4 above as an example. A blank version is available in Appendix B.



Execute

Execute your plan, analyse the collected data and evaluate the success of your program.

Now that you have planned and designed your program and evaluation, it's time to put it into action.

This section will guide you through the execute phase of your evaluation. First, you will administer the evaluation. This involves recruiting participants and collecting the data. Then, you will analyse the data you collect. This section provides advice about how to analyse quantitative and qualitative data and suggests useful tools.



Recruit participants: invite people to take part in your evaluation

Consider how you will invite participants to take part in the evaluation. Will you send an invitation by email? Put a post on social media? Ask participants in person or on the phone?

Consent is mandatory. You must get consent from the participants. Participating in the evaluation is voluntary. Thus, you must ask them if they choose to take part. If they agree, then you must collect evidence of their consent. Get them to sign a form or tick a box online. If over the phone or in person, then get them to verbally state their consent. Talk to your organisation to check any legal requirements and standard procedures. Follow the guidelines for consent in the [National Statement on Ethical Conduct](#). Keep in mind that many people may not want to participate. That's normal. Make them feel comfortable and safe to decline your invitation.

Before you get consent, make sure that you give participants information about the evaluation. They must be informed and understand the evaluation before they consent. Tell them what the evaluation is about and how you will collect and use the data. Inform them about the benefits of taking part in the evaluation and any potential risks. Tell them if you will keep their data confidential or anonymous, and how.



If you use young people under the age of 18 in your evaluation, you **must** get the consent of their carers or guardians. Always get a **signature** as proof of consent from carers or guardians.

A few examples:

1. If you are using an online survey for your evaluation, then include the basic information on the first page. Then, get participants to tick a box saying that they read it and agree to take part. Make the tick box a requirement to enter the online survey. If they don't consent, they don't do the survey.
2. If you are doing a focus group, give participants a form to read and sign beforehand. The form should include all the information about the evaluation and a place to agree and sign.
3. If your participants are under 18, send an email to their carers or guardians. Include an electronic form (e.g. pdf document) that they can sign and email back to you. Or, send them a link to an online form to electronically sign and submit.

Appendix C contains templates of participant information and consent forms (adult and child). Use them to create your own. Adapt them to suit your needs and your evaluation.

Collect the data: gather, document and record information



Once you recruit and get consent from your participants, you can start to collect the data.

Keep good records. Document detailed information in spreadsheets, notebooks and documents. Record dates, times and locations. Take note of any issues or problems that arise. Write down any changes to the procedures and methods. Be as accurate and precise as you can.

Keep in mind that you will need to report on your evaluation and findings. You want to make reporting as easy as possible. Precise and detailed data collection and record-keeping will make reporting easier. It will also make it easier for you, or others, to replicate your evaluation. Replication is important to see if the same findings occur in similar programs.



If you are using a survey instrument, there are a plethora of free online survey tools that you can use to build and administer your survey. Do a quick internet search for 'free survey tools' to explore the options that suit your needs.



Carefully manage the data collection. Make sure you adhere to the protocols you developed around data storage, access and confidentiality (in the 'Design' section on p. 39).



Analyse the data: examine, summarise and explain the data

Once you have collected all the data from your participants, it's time to analyse it. Summarising and analysing data allows you to interpret information and give it meaning. This section provides practical guidance and helpful resources to analyse quantitative and qualitative data.



The [Generic Learning Outcomes](#) website contains tools and spreadsheet templates to analyse quantitative and qualitative data. The tools and templates link to the five outcome categories presented in Table 1 on p. 17.

Analysing quantitative data

Quantitative data are numbers that can be analysed using statistics. The results are then displayed as tables, charts, histograms, and graphs.

You don't need to be an expert statistician to analyse your data. In fact, most online survey platforms have built-in tools that analyse your data for you. There are two basic elements to quantitative analysis: descriptive statistics and data visualisation.

DESCRIPTIVE STATISTICS

Descriptive statistics do just that—they describe your data. At the most basic level, descriptive statistics include counts and percentages.

Example:

Say 20 people replied to your survey. You count 15 'yes' responses and 5 'no' responses to a question. From this, you calculate the percentages: 75% of respondents said 'yes' and 25% said 'no'. If you are using an online survey platform, the built-in analysis tools will do this for you.

Counts and percentages are the simplest and most common way to represent data. They are an excellent first step in analysing your data. You can use them on their own, but if you can, consider pairing them with central tendency and dispersion measures (see the '*Statistically speaking*' box).



Using an online survey? Some platforms, such as [Survey Monkey](#), allow you to dig deeper into your data using customisable analysis features. For example, [filters](#) let you break down your survey results to focus on specific parts of your data. [Compare rules](#) let you choose two or more answer options from a single question and view them side-by-side. Explore the options that are available on the platforms that you decided to use for your evaluation.

STATISTICALLY SPEAKING

More descriptive statistics

It is also useful to summarise where and how spread out the data are. These measures are called central tendency and dispersion. Calculate them and pair them with your counts and percentages. It's easy to do using Microsoft Excel. There are [support sites](#) to help you (specific links are included below).

Central tendency is a single number that best represents your data. There are three measures of central tendency:

The [mean](#), or 'average', is the most common measure. Add all the numbers and then divide by the number of values.

The [mode](#) is the most frequently occurring number in the data.

The [median](#) is the midpoint in a set of ordered numbers.

Dispersion is a single number that gives you a sense of how spread out your data are.

The [standard deviation](#) is the most common measure of dispersion. It tells you how much your data disperse (stray) from the mean. A large standard deviation means that the data differ a lot, while a small one means they tend to be similar.

Inferential statistics: beyond this guide

Inferential statistics are more complex and are **beyond the scope of this guide**. Rather than describe the data, they make inferences, predictions or draw conclusions about the data. There are many kinds of inferential statistics (e.g., t-tests, ANOVA and regression).

Inferential statistics require more powerful tools beyond Excel, such as: [SAS](#), [SPSS](#), [STATA](#), [R](#), and [Minitab](#). These tools use coding languages and are more complicated to use. If you have the expertise or the will to learn, they are useful and powerful tools.

VISUALISE YOUR DATA

Make the trends, patterns, and findings in your data clear and easy to interpret. Visualise the results of your analysis using tables, charts and graphs.

Organise the quantitative results in tables. Keep your tables simple and clear. Avoid including too much information. Decide what information is most important (based on your priorities) and focus on those results. If you are using an online survey platform, the built-in analysis tools will create basic tables of counts and percentages for you.

Represent your data in charts or graphs. Common charts and graphs include bar charts, pie charts, scatter charts, line graphs or treemaps.

- Bar charts are the most common type of data visualisation. They show the values of different categories of data as rectangular bars with different lengths. You can make bar charts that are vertical, horizontal, or even stacked on top of one another.
- Pie charts are a circular graphic that is divided into slices to show proportion. They let you easily compare data in the form of slices.
- Scatter charts use dots to represent values for two different variables. You can use them to show relationships between variables.
- Line graphs use lines to connect data points (dots) in a 'dot-to-dot' fashion. You can use them to track how data changes over time, especially when there are many data points.
- Treemaps display hierarchical data using nested rectangles of sizes proportional to the corresponding data value.

There are many more ways you can visualise data. Select the option that best represents the data that you want to visualise. Excel has many options—take a look at their [support site](#) for ideas, help and guidance. If you are using an online survey platform, most of them have built-in options of charts and graphs for you to select from.

ANALYSING QUALITATIVE DATA

Qualitative data is any information that is not numerical. It can be in the form of text, images, audio and video recordings. There are many ways to analyse qualitative data, including word clouds, tagging and sentiment analysis.

- **Word clouds** are a common way to analyse textual qualitative data, like open-ended responses. They let you quickly interpret open-ended responses with a visual display of the most frequently used words. You can use one of the many free online word cloud tools. Many online survey platforms have word cloud tools built into their analysis tools.
- **Tagging** is another common technique where you go through the data, identify categories and tag different pieces of data to them. As you explore the data, you might identify new categories as new ideas or topics emerge. The tagging process is iterative, meaning that the categories and tags evolve and change as you analyse the data.

You can tag the data based on themes, topics, ideas or keywords. You can do this manually if there is not too much data. Use highlighters, cards and sticky-notes to organise and tag the data. You can also tag data in an Excel spreadsheet (see [Generic Learning Outcomes](#) spreadsheet template). Another way is to use computer programs for qualitative data analysis.

They are especially helpful if you have a lot of data. Examples of computer programs include [NVivo](#), [HyperRESEARCH](#), [ATLAS.ti](#). These programs help you organise, categorise and tag your data. They also have features that allow you to visualise the ideas, topics and themes into mind maps, treemaps and word clouds.

- **Sentiment analysis** is an emerging technique used to summarise the feelings behind written responses. Sentiment analysis uses natural language processing and machine learning to determine whether a response is positive, negative or neutral. It provides a high-level summary of how people feel in a chart or table. You can then filter a question by sentiment to dig further into the data and get deeper insights. Some online survey platforms (e.g., Survey Monkey, Qualtrics) have built-in sentiment analysis tools to help you analyse qualitative data.

ANALYSING MIXED METHODS DATA

If you chose to use a mixed-methods approach, then you will apply the analysis for both types of data. Analyse your quantitative data using statistics. Analyse your qualitative data using the techniques suggested above.

Triangulate the results. Remember, triangulate means to use more than one method and compare the results to see if they support the same findings (p. 34). Do the quantitative and qualitative findings support each other? If so, how? If not, how do they differ?

Quantitative and qualitative data give a different perspective and level of detail. The qualitative analysis can elaborate on or give context to the numerical quantitative results. Qualitative data provide more nuanced explanations than the numbers can. What does the qualitative data tell you about the quantitative data? Does it reveal any new insights that are not obvious from the quantitative results?

Example:

Take the [STEM Women](#) online directory as an example again (previously seen in the Define section p. 13). A posttest survey for the women who created profiles in the directory asks Likert scale questions (quantitative), such as:

Having a profile on STEM Women has made me feel more visible
[Strongly agree – Agree – Neither – Disagree - Strongly disagree]

The quantitative analysis of this question would involve counting how many respondents selected each of the answer options. The percentages could be calculated (e.g., 57% strongly agreed; 24% agreed). The analysis could also involve calculating the mean of the responses and the standard deviation.

What we don't know from this data and analysis is the context. Another question (qualitative) asks:

How has the database helped you build your professional network?

Here is an example of a response: "The database has really helped raise my professional profile. In just a few months since I joined, I have been invited to give a plenary at a conference, host a panel at a symposium and attend a high-profile networking event. Not only have these events been valuable in themselves, but they have also led to further connections and opportunities."

The responses could be analysed using word clouds, tagging or sentiment analysis (e.g., 62% of responses were positive; 43% contained the tag 'networking').

The qualitative analysis provides more detail about what, how, when and how much. It supports and complements the quantitative data.

YOUR TURN

Decide how you will analyse your data. What kind of data did you collect? What analysis techniques will best suit your data?

For quantitative data, calculate the frequency, central tendency and dispersion of your data. Visualise it using tables, graphs and charts. For qualitative data, explore and label your data into categories, themes or ideas. Do this manually or use a computer program. If you have more than one kind of data source, triangulate your results. Compare and contrast. Do the findings from different data and analyses support each other? How? If not, how do they differ?



Share

Share your findings publicly so that people can know if the program was effective and can work to improve future programs.

Sharing is the last step in the evaluation process. It is also a crucial step. Sharing your findings means that people can know if the program was effective. They can learn what worked, what didn't and how to improve future programs. Sharing your findings openly makes you accountable to your funders, partners and other stakeholders. It's also a way to celebrate your achievements.

This section will guide you through the sharing phase of the evaluation process. It provides advice on how to produce a report about your evaluation. Then, it offers ideas on how to disseminate your report and findings for maximum impact.

Report the results: organise, summarise and present the findings

First, you need to produce a report of your program evaluation.

This section highlights the importance of transparency in reporting your findings. It then prompts you to consider the evaluation audience—those who are interested in the findings of your evaluation. It also provides advice on how to organise your report.



It helps to look at evaluation reports from other programs. Take a look at the reports of STEM gender equity programs in Australia at [womeninstem.org](https://www.womeninstem.org).

Transparency is key

Be transparent and honest. Do not cherry-pick your results. Present everything—the good, the bad and the ugly. Transparency in evaluation is crucial.

The purpose of the evaluation is to understand what works or does not work, how, for whom, and why. If you only report the good, then others cannot assess the program or learn from it. Reporting negative results is as important as reporting the positive. Maybe even **more** important. Negative results can help us identify what doesn't work. Then, we can either avoid those things or try to improve them.

Negative results don't reflect poorly on your program. Instead, they show leadership and initiative—that you tried and tested something new.

Consider your audience

Keep your evaluation audience in mind. You defined your evaluation audience in the Define section (p. 14). Who is interested in the evaluation findings? Who will be reading this report? Who will use the findings and insights? Remember: Evaluation audiences can be program designers, facilitators and others in your organisation. They may also be program partners, funders and Government agencies. The local community may also be interested in the evaluation findings.



Also, consider how the findings and report will be used? Will the report inform future program development and delivery? Will you use it to secure future funding? Will the findings inform decisions relating to gender equity in education and workplaces?

Keep your evaluation audiences in mind as you produce your report.

Organise your report

A well-organised report will be easy for your audience to read and understand. There are many ways to organise your report. Be sure to include the following sections:



Executive summary. Give a brief overview or preview of the entire report. Present the key points and findings. Keep the summary short—1 to 2 paragraphs and no longer than one page.

Background of program and evaluation. Briefly describe the context and background of the program. What STEM equity issue did the program address? Why is that important? Describe the program itself. What were the program audience, goals and activities? Give an overview of the evaluation. What were the evaluation priorities and key questions? Use all the work that you did in the 'Plan' section to help you write this section.

Method. Describe the design of the evaluation. What design approach did you use? How did you recruit participants? What data did you collect and how? How did you analyse the data? Explain the approaches and techniques you used. Use all the work that you did in the 'Design' section to help you. Provide enough detail that someone could replicate your evaluation by reading this section.

Results and discussion. Present the results of the analysis in a logical way. Present all results—positive and negative (see ‘Transparency is key’ on p. 50). Include tables, graphs, charts. Keep your results clear and simple. What are the most important results that you want people to know about? Make them as obvious as possible. Compare and contrast between data sources. Explain your findings. What do the results mean?

Conclusions. Summarise the key findings. Relate them to the goals of the program. What can you conclude from the findings? What can we learn from them? How are they important to the reader? Be clear, specific and brief.

Recommendations. Make recommendations based on your evaluation. What should we do with the insights from the findings? How can we build on the positive results? How can we address and improve the negative results? Make your recommendations action-oriented and feasible. Arrange them in order of importance. Keep them brief.



Make your conclusions and recommendations impactful. Highlight what was effective. Suggest possibilities for revisions and future directions. Keep them action-oriented.



Make sure that the report is culturally, socially and ethically sensitive. Consider how the findings will affect participants and stakeholders of different cultures, groups and backgrounds. Be respectful of their values, needs, beliefs and basic assumptions. Consult them before disseminating the report. Use the [AIATSIIS guidelines](#). Adhere to your protocols around data participants’ confidentiality and privacy (p. 39).



Be fair to the range of possible interpretations that can come from your findings. Consider sharing a draft of your report with colleagues and stakeholders. Ask for feedback and comments on your interpretations, conclusions and recommendations.

YOUR TURN

Write your evaluation report. Always keep your evaluation audience in mind. Who will be reading your report? Who is interested in your findings? Organise your report as suggested above. Be clear, specific, honest and transparent. Make your conclusions and recommendations impactful and action-oriented.

Disseminate your report: share your evaluation far and wide



Share your evaluation report with key stakeholders and the broader public. Get maximum impact from your evaluation.

Make it public

Spread the word. Share your evaluation publicly. Your STEM gender equity program is a subject of national interest. It is part of a bigger effort to create change. The public has a vested interest in that change. They want to know how your program contributes to it. Key reports and policy documents, such as *Advancing Women in STEM* and the *Women in STEM Decadal Plan*, stress the value of publicly sharing STEM gender equity program evaluations [1, 3].

Get maximum impact—share far and wide. Publish the report on your organisation's website. Post about the results on social media. Write an article for an online news outlet. Talk about the findings on the radio or a podcast. Make a flyer to share around the local community.



Remember to be culturally sensitive when sharing your findings publicly. Avoid being unduly promotive. Sharing should try to advance STEM equity practices while maintaining integrity.

Share with key stakeholders

Share your report with key stakeholders—your evaluation audience. Share the report with key staff in your organisation. Your findings can inform future or ongoing programs. It also gives credibility to your program and justifies the resources you invested.

Send the report to your funders (if you have any). They are interested in how you used their investment and what the program achieved. The report shows that you are accountable for the expenditure, decisions and outcomes of the program. Your findings will help funders determine what programs should be scaled up. Your evaluation may help you secure funding for future programs. Even if your findings were not all positive, your report shows how you would improve it next time.

Disseminate the report to program partners or other stakeholders who might be interested. Present your findings at a conference, symposium or meeting. Sharing your evaluation demonstrates transparency and accountability. Your evaluation can inform best practice among other organisations. It can also uncover opportunities to collaborate or share resources.

Consider publishing

Turn your report into a paper for an academic journal. The academic community also has a vested interest in your findings. Your evaluation could contribute to the research about STEM gender equity issues.

If you choose to submit your evaluation as a journal article, you will need to rewrite it as an article. This can take work, but it can be valuable and rewarding. Many journals publish articles about STEM gender equity issues. Here is a list of potential journals to consider:

- [Gender and Society](#)
- [Journal of Gender Studies](#)
- [Journal of Women, Politics & Policy](#)
- [Gender & Development](#)
- [Gender Issues](#)
- [International Journal of STEM Education](#)
- [Journal of STEM Education Research](#)
- [Australian Journal of Education](#)

REFERENCES

1. Australian Government. (2019). *Advancing Women in STEM*. Department of Industry, Innovation and Science. <https://www.industry.gov.au/sites/default/files/March%202020/document/advancing-women-in-stem-strategy.pdf>
2. Corrigan, D., & Aikens, K. (2019). *Barriers to participation in engineering and the value of interventions to improve diversity*. Melbourne, Australia: Monash University. <https://educationfutures.monash.edu/news-events/barriers-to-participation-in-engineering-a-report-for-the-engineering-for-australia-taskforce>
3. Australian Academy of Science. (2019). *Women in STEM Decadal Plan*. <https://www.science.org.au/files/userfiles/support/reports-and-plans/2019/gender-diversity-stem/women-in-STEM-decadal-plan-final.pdf>
4. McKinnon M. (2020). *The absence of evidence of the effectiveness of Australian gender equity in STEM initiatives*. Australian Journal of Social Issues. Advanced online publication. <https://doi.org/10.1002/ajs4.142>
5. Phillips, T. et al. (2014). *User's Guide for Evaluating Learning Outcomes in Citizen Science*. Cornell Lab of Ornithology. https://www.birds.cornell.edu/citizenscience/wp-content/uploads/2018/10/USERS-GUIDE_linked.pdf
6. AusAID. (2012). *Impact Evaluation: A discussion paper for AusAID practitioners*. DFAT, Office of Development Effectiveness. <https://www.dfat.gov.au/sites/default/files/impact-evaluation-discussion-paper.pdf>
7. OECD. (2001). *Evaluation feedback for effective learning and accountability*. OECD Publications Service. <https://www.oecd.org/dac/evaluation/2667326.pdf>
8. Neresini, F., & Pellegrini, G. (2014). Publics and their participation in science and technology: changing roles, blurring boundaries. In M. Bucchi & B. Trench (Eds.), *Routledge Handbook of Public Communication of Science and Technology* (Second ed., pp. 231-245). Abingdon, Oxon; New York, NY: Routledge.
9. O'Shea, S., Howard, S., Lipscombe, K., Buckley-Walker, K., & Firth, W. (2019). *UOW Framework for Evaluation of Equity Programs*. University of Wollongong: Wollongong: Australia. <https://www.uow.edu.au/about/learning-teaching/hepppl/>
10. W.K. Kellogg Foundation. (2004). *Using Logic Models to Bring Together Planning, Evaluation, and Action: Logic Model Development Guide*. Michigan: W.K. Kellogg Foundation. <https://www.aacu.org/sites/default/files/LogicModel.pdf>
11. Biggs, J.B., & Tang, C. (2011). *Teaching for quality learning at university: What the student does*. Fourth ed. McGraw-Hill Education: London, UK.
12. Arts Council England. (2018). *Generic Learning Outcomes*. <https://www.artscouncil.org.uk/measuring-outcomes/generic-learning-outcomes#section-1>
13. University of Minnesota. (2019). *Evaluation Design*. <https://cyfar.org/evaluation-design>
14. McAlpine, L. (2016). *Why might you use narrative methodology? A story about narrative*. Eesti Haridusteaduste Ajakiri. Estonian Journal of Education. 4(1): 32-57. <https://doi.org/10.12697/eha.2016.4.1.02b>
15. Australian Government. (2015). *Australian Government Guidelines on the Recognition of Sex and Gender*. <https://www.ag.gov.au/sites/default/files/2020-03/AustralianGovernmentGuidelinesontheRecognitionofSexandGender.pdf>

GLOSSARY

TERM	DEFINITION
Evaluation	The systemic collection of data to assess the strengths and weaknesses of program. Evaluation also seeks to establish whether, and to what extent, an intervention made a difference by measuring the positive or negative changes produced by an intervention. The changes can be direct or indirect, intended or unintended. Evaluation aims to answer questions about what works or does not work, how, for whom, and why [5, 6, 7].
Equity	The practices and ways of thinking that assist in working towards equality, including ensuring individuals or groups of individuals are given opportunities and resources that are proportional to their needs. Equity differs from equality in that it acknowledges that under-represented groups do not start from the same point, may face different systemic barriers, and therefore may require additional support to overcome these barriers [3].
Gender	Gender is part of a person's personal and social identity. It refers to the way a person feels, presents and is recognised within the community. A person's gender may be reflected in outward social markers, including their name, outward appearance, mannerisms and dress. A person's sex and gender may not necessarily be the same. An individual's gender may or may not correspond with their sex assigned at birth, and some people may identify as neither exclusively male nor female [15].
STEM	Refers to science, technology, engineering and mathematics, and is used here as an umbrella term for scientific and technical fields. Medical research is included in the definition of STEM used here; however, the practice of health professionals is excluded. STEM skills are those taught and used in STEM disciplines, but the term is used more broadly to encompass the creativity and ways of thinking necessary to promote innovation and problem-solving in any discipline or industry [3].
Woman/Women	Anyone who identifies as a woman, including cisgender (personal gender identity corresponds with sex assigned at birth), transgender (personal gender identity does not correspond with sex assigned at birth), non-binary and intersex persons who identify as a woman (or girl) [3].

The background features a collection of light gray icons on a darker gray background. These icons include: a hand pointing to a target, a hand with the index finger pointing up, two gears, a large eye, a key, and a notebook with a pen resting on it. The word "Appendices" is centered in the lower half of the page.

Appendices

Appendix B: Data management plan

Use the table below to develop a data management plan. Refer to p. 39 for details and an example.

DATA TYPE	HOW IS DATA STORED SECURELY	HOW IS ACCESS RESTRICTED	HOW LONG WILL DATA BE KEPT
Hard Copies			
Electronic Copies			
Audio/Visual			

Appendix C: Participant Consent Form Templates

For adults (18 years and over)

I would like to invite you to take part in the evaluation of the program [name of program]. I am reaching out to you because [reason: you are attending the program; you facilitated the workshop]. I ask that you please read the information below and ask any questions before you agree to take part.

What is the evaluation about? The purpose of this evaluation is to [insert the relevant priorities you identified in the planning phase]. The evaluation will answer the following questions:

- [insert the key question you developed in the planning phase]

Do I have to take part? Participation in this study is voluntary. Your decision to take part will not affect your current or future relationship with [name of organisation or program]. If you choose to take part, you are free to not answer any question or withdraw at any time.

What do I need to do? If you choose to take part in this evaluation, I will ask you to [describe what they will do: complete a survey; complete a task; participate in a focus-group; take part in an interview; be observed]. This will happen on [date(s); time(s); location(s)]. It will take approximately [amount of time: 10 minutes, 1 hour].

What are the possible benefits of participating? Are there any risks? The information from this evaluation will help [insert the relevant priorities of your evaluation].

I do not expect there will be any type of risk associated with participating in this evaluation. [If there are risks, outline them here].

What will happen to information about me? I will use the information collected from you to evaluate [insert the relevant priorities of your evaluation]. I will store the information in [describe your data management plan you developed in the design phase]. The information will be kept for [refer to your data management plan]. I will publish the findings of the evaluation [a report; an article; a presentation]

Will I be identified? I will use [pseudonyms, ID numbers] to protect your identity OR I will not collect any identifying information about you.

Contacts and questions: If you have any questions about participating in the evaluation, please contact [name of contact] at [contact details: phone; email].

Consent:

- I have read the information and agree to take part in the evaluation.

Signature: _____

Date: _____

For children (under 18 years)

I would like to invite your child to take part in the evaluation of the program [name of program]. I am reaching out to you because [reason: your child is attending the program; your child used the resources]. I ask that you please read the information below and ask any questions before you agree for your child to take part.

What is the evaluation about? The purpose of this evaluation is to [insert the relevant priorities you identified in the planning phase]. The evaluation will answer the following questions:

- [insert the key question you developed in the planning phase]

Does my child have to take part? Participation in this study is voluntary. You and your child's decision to take part will not affect your current or future relationship with [name of organisation or program]. If you and your child choose to take part, they are free to not answer any question or withdraw at any time.

What do they need to do? If you and your child choose to take part in this evaluation, I will ask them to [describe what they will do: complete a survey; complete a task; participate in a focus-group; take part in an interview; be observed]. This will happen on [date(s); time(s); location(s)]. It will take approximately [amount of time: 10 minutes, 1 hour].

What are the possible benefits of participating? Are there any risks? The information from this evaluation will help [insert the relevant priorities of your evaluation].

I do not expect there will be any type of risk associated with participating in this evaluation. [If there are risks, outline them here].

What will happen to information about my child? I will use the information collected from your child to evaluate [insert the relevant priorities of your evaluation]. I will store the information in [describe your data management plan you developed in the design phase]. The information will be kept for [refer to your data management plan]. I will publish the findings of the evaluation [a report; an article; a presentation]

Will they be identified? I will use [pseudonyms, ID numbers] to protect your child's identity OR I will not collect any identifying information about your child.

Contacts and questions: If you have any questions about your child's participation in the evaluation, please contact [name of contact] at [contact details: phone; email].

Consent:

- I have read the information and provide permission for my child to take part in the evaluation.

Signature of carer/guardian : _____

Date: _____

Appendix D: Sample Focus-group protocol

This appendix provides an example of a focus group protocol.

Facilitator: Program leader

Recorder: Program administrative support

Invited participants: Program designer (1) and facilitators (3)

Time: 1 hour

10 mins: general discussion

- What are your general thoughts about the program?
- Did the program address girls' attitudes to STEM study and careers?
- Did the program target the right audience?
- Did the program run smoothly?
- Was the program adequately resourced?

15 mins: strengths

- As a group, create a mind map the strengths of the program on the whiteboard.
- What are you satisfied about? Why is that?

15 mins: weaknesses

- As a group, create a mind map the weaknesses of the program on the whiteboard.
- What are you dissatisfied about? Why is that?

20 mins: reflection and suggestions

- Compare and contrast the two mind maps. Do any concepts/ ideas relate to or influence each other?
- What aspects of the program could change? How?
- Is this particular program the best option to achieve the program goals?

Data will include:

- Written notes of the discussions by the recorder
- Photographs of the mind maps on the whiteboard

Appendix E: Checklist

Here is a checklist summary of the evaluation process covered in this guide. Use this checklist to make sure you have covered all five essential steps of quality program evaluation.

Step 1: DEFINE

- Define the problem: What issues does your program address? (p. 12)
- Define audiences: Who is the program for? Who is the evaluation for? (p. 13)
- Define program goals: What change do you want to create? (p. 15)

Step 2: PLAN

- Select program activities: What will you do to achieve your goals? (p. 19)
- Identify evaluation priorities, key questions and indicators: how will you measure it? (p. 22)

Step 3: DESIGN

- Choose the evaluation approach: How will you carry out your evaluation? (p. 28)
- Decide on the data collection method: What data will you collect and how? (p. 30)
- Create your instrument: Use an existing one; if you must, modify or create your own (p. 35)
- Run a pilot: Test your method and instrument to make sure it works properly (p. 39)
- Develop a data management plan: Where and how will you safeguard data? (p. 39)

Step 4: EXECUTE

- Recruit your participants: How will you invite participants and get their consent? (p. 42)
- Collect the data: Gather, document and record information (p. 43)
- Analyse your data: What kind of data do you have? Examine, summarise and explain it (p. 44)

Step 5: SHARE

- Report honestly and transparently: Report the positive and the negative results (p. 50)
- Consider your audience: Who will read your report? Who is interested in the findings? (p. 51)
- Organise your report using the structure outlined on p. 51
- Share the evaluation publicly and with key stakeholders (p. 53)
- Consider publishing: Write a paper and submit to an academic journal (p. 54)



An Australian Government Initiative

Women in STEM Ambassador