EVALUATING STEM EQUITY PROGRAMS

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 and Resources.

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

December 2020

Cite as: Kingsley, I. (2020). *Evaluating STEM Equity Programs: A guide to effective program evaluation*. Office of the Women in STEM Ambassador. https://womeninstem.org.au/national-evaluation-guide

CONTENTS

Introduction

Why is evaluation of STEM equity programs important?

"If you can't measure it, you can't improve it." — (p. 23)

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

Australia has hundreds of programs to tackle STEM 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 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 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]S ou ⁶ † y ! ' e f c]

Evaluation basics

This guide provides advice on how to evaluate your STEM 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 and is embedded within the program from the very beginning.

Four areas to evaluate

Evaluations should assess four main program areas: Design, E f ciency, Outcomes and impacts, and Lessons learned. This guide mainly focuses on evaluation of ; 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 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.

E f ciency. 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 e f ciency at any time or continuously throughout the program.

Lessons learned. Re f ect 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

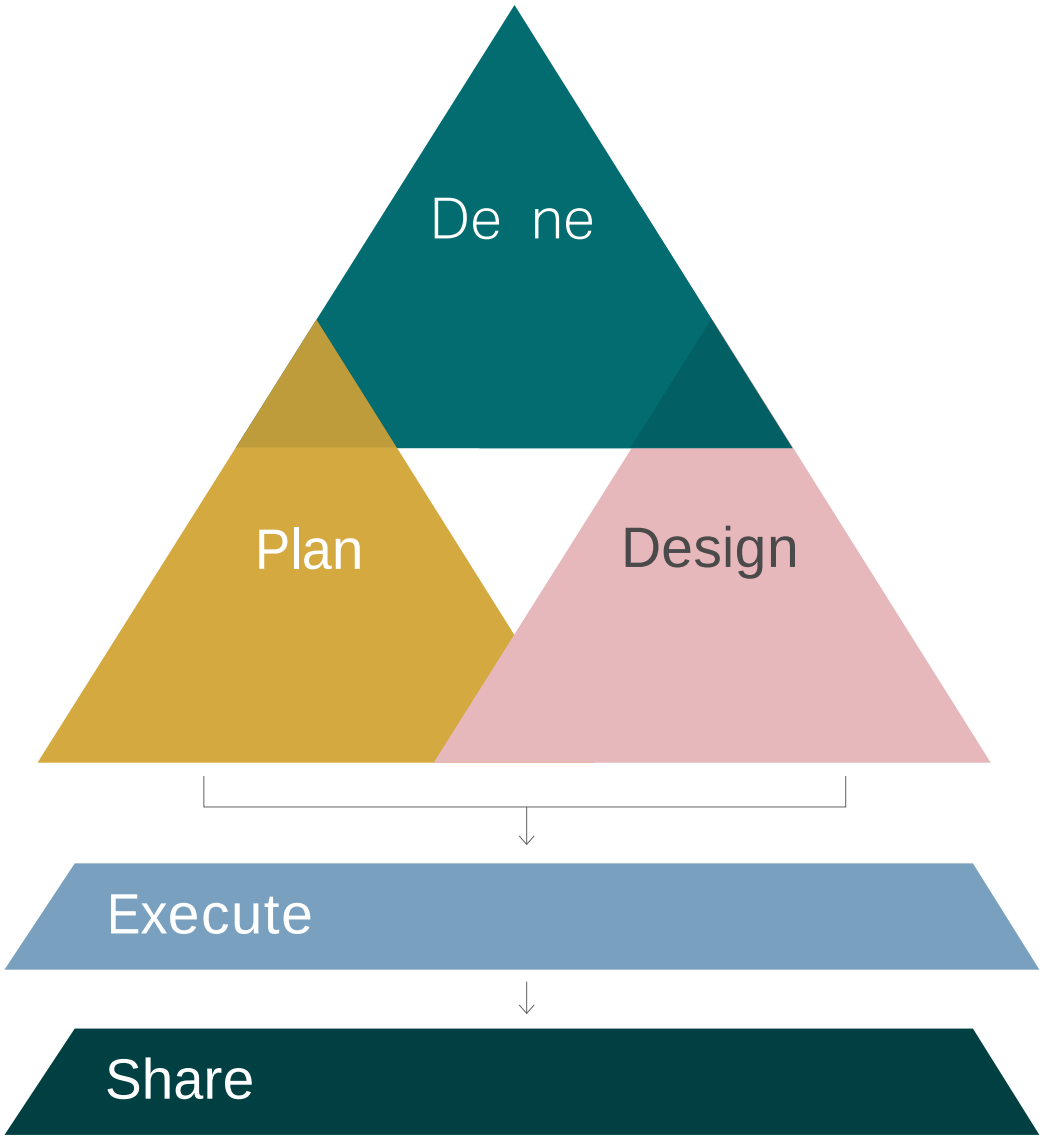
This guide o f ers practical advice to help you—STEM equity program leaders and evaluators—evaluate your program. It does not cover every nuanced aspect of evaluation. Instead, it gives in-depth guidance on targeted concepts. This guide focuses on evaluating program outcomes and impacts.

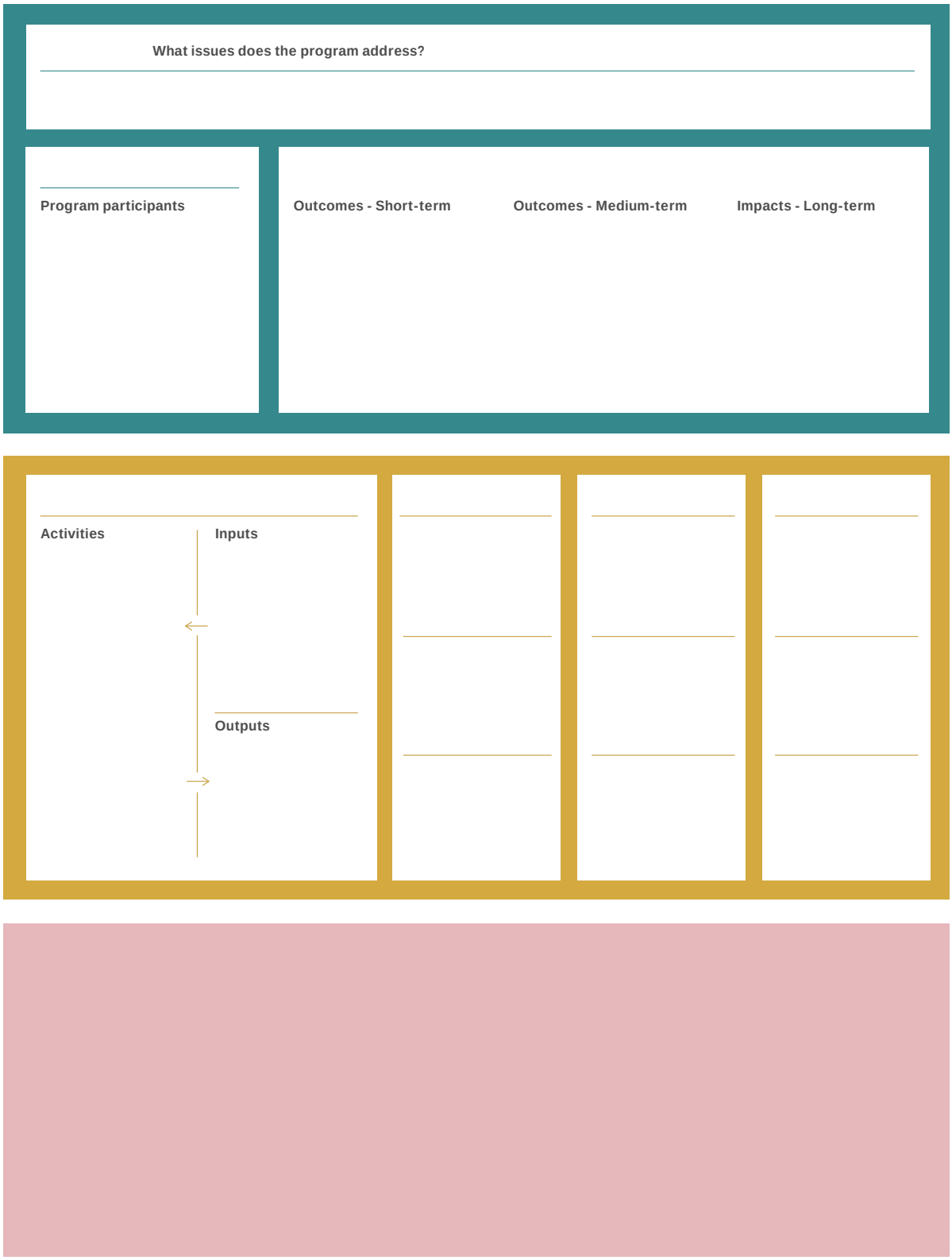
Program evaluation does not need to be daunting, di f cult, expensive or time-consuming. This guide breaks it down into a simple, 5-step approach that's easy to apply to your program. It shows you how to embed evaluation within your program from the very beginning.

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. Use this guide with the STEM Equity Evaluation Portal. It is an easy-to-use online tool that helps you plan and report your evaluation using the concepts in this guide.

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.





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

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

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 inss

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: earch (A' W

Defne

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].

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.

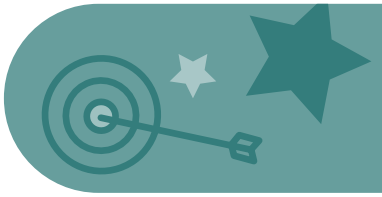
Defining who the evaluation is for will help you determine its priorities (see p. 22). It will also help you tailor your report and how you disseminate the findings (see p. 51).

GOVERNMENT: A KEY AUDIENCE

Government is a key 'evaluation audience.' The Australian Government

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.

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 girls, women, non-binary and other marginalised groups participation in STEM education or careers. Unless STEM 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 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. Use the STEM Equity Evaluation Portal to help you define your goals. It contains an extensive list of equity program goals for you to select from.

KNOWNS AND UNKNOWNS

There is a wide variety of STEM equity programs. These programs have many different intended *outcomes* that all contribute to shared future *impacts* on 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 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 unknows (impacts) is not easy to establish.

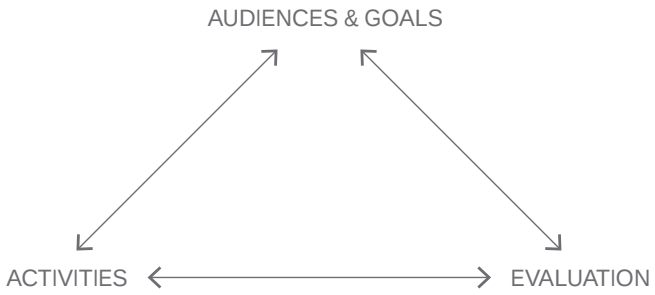
Consult your potential audiences or relevant stakeholders when defining the outcomes and impacts. Consultation is a continuous process, but it's most important in the planning phase. It will help you make sure that the outcomes and impacts are appropriate and sensitive to the groups who may be affected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

OUTCOME CATEGORY	DEFINITION
Enjoyment & Inspiration	• Having fun • Innovative thoughts • Exploration, experimentation and making • Being inspired
Attitudes and values	• 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	• Knowing about something • Learning facts or information • Making sense of something • Deepening understanding • Making links and relationships between things
Skills	• Knowing how to do something • Being able to do new things • Intellectual skills •

Plan

Plan the program activities and evaluation so that they align with the audiences

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 together[1]. Think about the goal that

- What will you do to achieve it?

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.

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

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.

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.

Evaluation: how will you know what is achieved?

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, confdence 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 fnd out if the program was successful in achieving these outcomes. It was also to fnd out what worked and what didn't to refne 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 " en l' s and ' m pp Ŏ > M ɒ ɹ ɕ Nʹ

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

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:

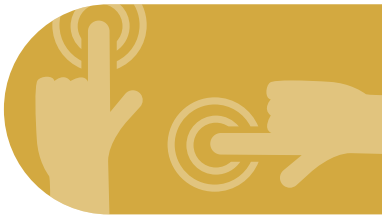
- Did the program run as planned?

Finally, a question related to lessons learned might ask:

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

Write the key evaluation questions in the STEM Equity Evaluation Portal or 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:

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

Percentages. The percentage of Indigenous students enrolled in VET and university STEM subjects; the percentage of 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 non-binary 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.

Use the STEM Equity Evaluation Portal—it contains lists of relevant indicators based on the goals that you define. Check the STEM Equity Monitor for ideas of indicators that could demonstrate change. It is an interactive national data report that shows the current state of STEM 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.

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.

Design

Design your program evaluation to determine how you will measure the achievement of your goals based on your plan from the previous step.

Table 2. Strengths and weaknesses of common evaluation designs

DESIGN APPROACHES	STRENGTHS	WEAKNESSES
Posttest *program posttest*	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 *pretest program posttest*	Measures change over time	Cannot account for other factors that might have influenced or caused the outcome
Time series *program (test 1, 2...) posttest*	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 *program pretest & posttest*	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 equity programs are delivered in complex and unstable environments. A lot of other factors come into play that might influence or cause the change. M p א ae p φ T ess pry я ' e "

m=es " g' cnte " A " es onm A lo. M ƽ ome inauseM

Data collection: what data will you collect and how?

Think about what data you want to collect based on your key

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 specifc 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 employees 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 yvi " ! ض l r

Case study. A detailed, up-close and in-depth look at a single ‘case’—a person,

Instrument: what tool will

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 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. 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**?

The STEM Equity Evaluation Portal takes all of the work out of choosing an instrument. It contains a collection of valid and reliable instruments selected by experts. Based on your program goals and indicators, the STEM Equity Evaluation Portal will recommend the appropriate instruments for your evaluation.

Selecting an existing instrument 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.

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?

- x number of STEM careers
- 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.

- 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.

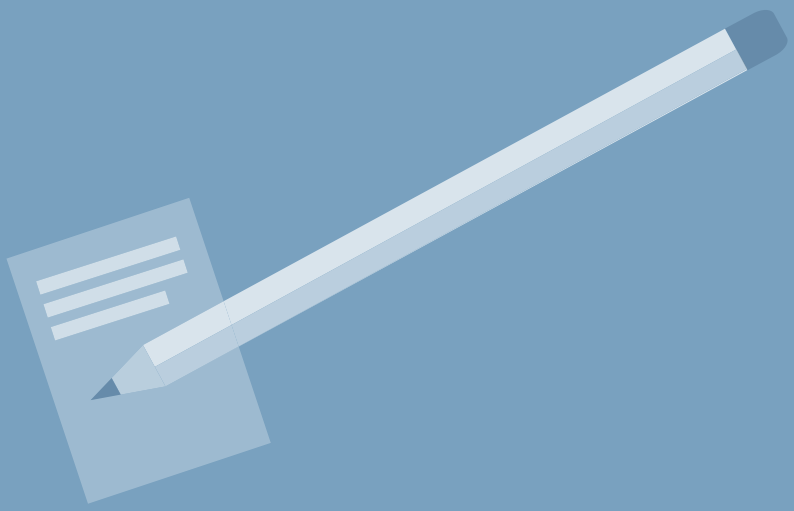
- your research to a 5-year old using no jargon
- x phenomenon using a metaphor
- a 1-minute video about your most recent research journal article

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

Create your instrument. Use the _____



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.

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.

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

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 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. 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.

Example:

Take the STEM Women

Share

future programs.

Sharing is the last step in the evaluation process. It is also a crucial

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 disseminate**

Report the results: organise,

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 evaluations published on the STEM Equity Evaluation Portal as examples.

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

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,

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 equity program is a subject of national interest. It is part of a bigger e fort 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 equity program evaluations [1, 3].

Get maximum impact—share far and wide. Publish your evaluation on the STEM Equity Evaluation Portal. It is a national database of equity programs to help us know what works and learn from each other to improve equity. 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 fndings on the radio or a podcast. Make a fyer to share around the local community.

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

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 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 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
- Equality, Diversity and Inclusion
- Journal of Women and Minorities in Science and Engineering
- Australian Journal of Social Issues
- Organization Studies

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/user files/support/reports-and-plans/2019/gender-diversity-stem/women-in-STEM-decadal-plan-final.pdf
4. McKinnon M. (2020). *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). . 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

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].
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 promot l' ciplimppis i s [3].

Appendices

Appendix A: Evaluation Planning Tool

Use this worksheet to help you define, plan and design your program evaluation.

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

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			

For children (under 18 years)

Appendix D: Sample Focus-group protocol

This appendix provides and 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 satisfed 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 dissatisfed 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)