Evaluating Behaviour Change Interventions: A Practical Guide

Lynette J. McLeod[1], Tanya M. Howard[1], Aaron B. Driver[1] and Donald W. Hine[1,2],
[1]University of New England, [2]University of Canterbury 2023

Evaluating Behaviour Change Interventions

A Practical Guide

Lynette J. McLeod
Tanya M. Howard
Aaron B. Driver
Donald W. Hine

University of New England
Armidale, Australia
University of Canterbury
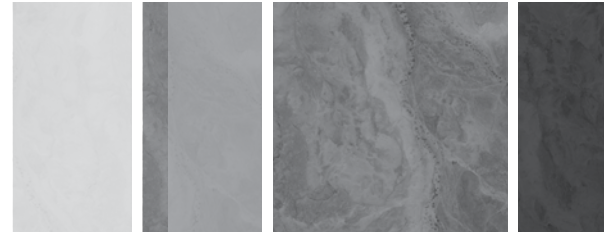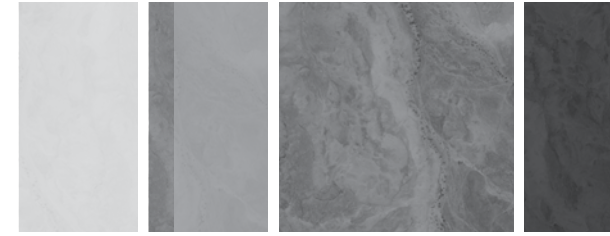Christchurch, New Zealand

table of contents

introduction

Scientists are very good at developing technologies and recommended best practices for managing sustainable land management problems. But these proposed solutions will fail unless the public – land managers and community members – are sufficiently empowered and motivated to modify behaviours and adopt new approaches.

Changing behaviour, and sustaining these changes over time, is a difficult process. Educating the public about adverse impacts, and providing information about control strategies, is rarely enough.

Behaviour change interventions require a more sophisticated approach, informed by behavioural sciences. Social psychology and behavioural economics have generated an array of intervention strategies and behaviour change techniques designed to increase audience understanding, engagement and, ultimately, adoption of desired behaviours.

'Designing Behaviour Change Interventions for Sustainable Land Management: A Practical Guide' (Hine, McLeod and Driver, 2022) provides a systematic approach for practitioners to develop new behaviour change interventions. There are four principles to follow:

1. Focus on behaviour.

2. Know your audience.

3. Match your interventions to the primary causes of behaviour.

4. Evaluate, review and reflect.

This guide focuses on the principle 'Evaluate, review and reflect'. It provides practitioners with a systematic approach for developing and implementing an appropriate evaluation plan for their intervention. A practical case study (involving domestic cat management) is used throughout, to demonstrate the theory content.

How to use this guide

This guide can be used in conjunction with 'Designing Behaviour Change Interventions for Sustainable Land Management: A Practical Guide'. It will enable scientists, policy makers and engagement specialists to better connect with their stakeholders and target audiences. Doing so should improve participation rates and effectiveness.

one

What is evaluation, and why is it important?

Evaluation is a process that critically examines an intervention. It involves collecting and analysing information about the intervention's characteristics, activities and outcomes in order to judge it, improve its effectiveness and/or inform future intervention decisions.

Evaluation is not only an accountability measuring stick to determine if an intervention works. It also improves the engagement and design process.

Evaluation can:

- provide insight into–
  - » purpose and goals for addressing the identified problem
  - » needs and wants of stakeholders and target audience
- improve intervention design and implementation of–
  - » stakeholder and audience engagement
  - » behaviour prioritisation
  - » behaviour change tool choice
  - » delivery plan and communications
  - » demonstrate the impact of the intervention via–
  - » changes in human behaviour over time
  - » improvement in the problem being addressed.

What makes a good evaluation?

A well planned and carefully executed evaluation will reap more benefits than one thrown together hastily and retrospectively. Though you may feel that you lack the time, resources and expertise to conduct an assessment, learning about evaluation early – along with careful planning – will help you navigate the process.

A good evaluation is:

- **Tailored to your program.** It should address the specific goals and objectives of your intervention and build on existing evaluation knowledge and resources.
- **Replicable, using rigorous methods.** The better your evaluation design, implementation and analysis, the more accurate its conclusions and the more confident others will be in its findings. Others should be able to conduct the same evaluation and get the same results.
- **Inclusive.** It should ensure all viewpoints are considered and that results are as complete and unbiased as possible.
- **Honest.** Results are likely to suggest that your program has strengths as well as limitations. Evidence that your intervention is not achieving all its objectives can be challenging to accept, but can also indicate where to best put your limited resources.

two

Developing an evaluation plan

Why should I develop a plan?

The earlier you develop a plan and begin to implement it, the better the outcomes will be.

Alter et al. (2017) offer six reasons why you should develop an evaluation plan:

1. guides you through each step of the evaluation process

2. informs you about the type of information to collect

3. prevents you from wasting time collecting information you don't need

4. helps you identify the best methods to use

5. allows you to set an appropriate time frame for evaluation

6. improves your intervention development practice.

Who should be involved?

You need to identify who needs to be involved in the evaluation and how to manage it. This means engaging with your stakeholders and better understanding them.

BetterEvaluation (2014) recommends that you determine:

- the best strategy to engage with your stakeholders (*e.g. formal or informal meetings, communication plan*)

- who has the authority to make decisions, provide advice or make recommendations about the evaluation (*e.g. Do you need to establish an advisory or steering group? How will decisions be made?*)

- the resources required (*e.g. funds, expertise*) and where they will be obtained

- who will conduct the evaluation (*e.g. internal staff, external consultant, peer review*)

- ethical and quality standards for consideration (*e.g. human research ethical guidelines, evaluation standards, cultural competency*)

- how to manage the evaluation processes (*e.g. establish Terms of Reference for the evaluation and, if required, contractual agreements or Memorandums of Understanding) for how different organisations will work together*)

- how to develop the evaluation processes (*e.g. choosing a suitable evaluation framework, creating an evaluation work plan*)

- how to monitor the evaluation process (e.g. critical reflection, expert or peer review).

Where do I start?

Before designing your evaluation plan, define what to evaluate, its purpose and its parameters.

To define your evaluation, you need to set out the objectives and develop a description of what needs to be evaluated.

The SMART framework provides a simple guide for setting realistic objectives. They should be:

- **S = Specific:** Be as clear and specific as possible about what you want to do.
  - » *Specific objective: At least 90% of new cat adopters will desex their cats by June 2022.*
  - » *Non-specific objective: To protect newly adopted cats from unwanted pregnancies.*

- **M = Measurable:** Can you measure whether you have achieved your objective?
  - » *Measurable objective: To increase full-time containment of pet cats by 50% over the next two years.*
  - » *Non-measurable objective: To fully contain pet cats.*

- **A = Achievable:** Is your goal achievable? Can you get it done in the time allocated, within your budget and with the available expertise?
  - » *Achievable objective: To reduce predation pressure from free-roaming pet cats on a nature reserve's wren population by 50% over the next five years.*
  - » *Non-achievable objective: To stop cats from killing wrens.*

- **R = Relevant:** Will achieving this objective contribute to delivering your intervention and its aims?
  - » *Relevant objective: Developing cat owner guidelines in suitable languages to improve compliance among people whose first language is not English.*
  - » *Irrelevant objective (or not relevant enough): To improve cat owner compliance among people whose first language is not English by teaching adults to read English.*

- **T = Time-based:** What is your available time frame? An end date can motivate, but your goal must be achievable within the set time.
  - » *Time-specific objective: To reduce by 50% the proportion of people within the council area who feed stray cats, by 30 June 2024.*
  - » *Non-time-specific objective: To reduce by 50% the proportion of people within the council area who feed stray cats.*

Table 1 Different types of evaluations (Rossi, Lipsey, & Freeman, 2004)

Type of evaluation	Purpose
Formative	
1. Needs assessment	Determines who needs the intervention, how great the need is and how to best meet it.
2. Process evaluation	Examines the intervention's theory and the process of developing and implementing it. Can be a continuous or one-time assessment.
Summative	
1. Outcome evaluation	Investigates the extent to which the intervention has achieved its outcomes, i.e. the observed behaviour of the targeted audience, that you expected to change.
2. Impact evaluation	Measures the effect of the intervention on its intended outcomes and any unintended side effects.
3. Efficiency analysis	Relates intervention costs to beneficial outcomes. Examples include cost-benefit and cost-effectiveness analysis.

What is the purpose of your evaluation? To define the purpose of your evaluation, consider how you will use its results. This will inform the type of evaluation you conduct and help frame its boundaries (BetterEvaluation, 2014).

The process of evaluation could:

- include diverse stakeholder perspectives
- build trust and legitimacy with stakeholders
- ensure accountability.

Your results could:

- inform decision making to improve the intervention design process (formative evaluation)
- inform decision making around selecting, continuing or ceasing an intervention (summative evaluation)
- contribute to the broader evidence base, to guide future policy and practice by other practitioners/organisations (formative or summative evaluation)
- be used to lobby and advocate, justify expenditure and demonstrate achievements (summative evaluation).

What will success look like? Define the criteria and standards for judging performance so your evaluation can achieve its objectives. Depending on the evaluation's purpose, you can derive these from:

- the goals and objectives of the intervention project
- other relevant standards, criteria or benchmarks
- stakeholders / public consultation / feedback.

What are the key questions to be asked? The purpose of your evaluation determines the key questions you will ask. These can take four formats (BetterEvaluation, 2014):

1. Descriptive questions – What has happened? What is the situation?
   For example: Where has the intervention been delivered? What changes have occurred for participants?

2. Causal questions – What caused or contributed to the results?
   For example: What were the outcomes and impacts of the intervention? What other factors contributed to these outcomes and impacts?

3. Synthesis questions – Is this the best option? In what ways could it be better?
   For example: Was the intervention cost-effective? What were its strengths and weaknesses?

4. Action questions – What action should you take?
   For example: Should the intervention be continued/scaled up? What changes should you make to it?

Figure 1 A simple example of a logic model for an intervention (after Alter et al., 2017)

Inputs	Outputs		Outcomes		
Intervention investments	Activities	Participants	Short term	Medium term	Long term
What is invested	What we do	Who we reach	Changes in awareness, attitudes, knowledge, skills etc	Changes in behaviour, relationships etc	Cultural shift in attitudes, decrease in negative impacts, increase in positive outcomes

A relatively easy way to understand how the intervention will work (i.e. the process or logic behind your intervention achieving its goals) is to create a logic model. This graphic representation maps the relationships between investments (inputs), activities and results (outputs, outcomes and impacts) (BetterEvaluation, 2014).

A simple approach is to use a 'results chain' model (pipeline model). Figure 1 shows an intervention as a series of boxes: inputs → outputs → outcomes (and impacts).

When defining what you are evaluating, it is essential to consider possible unintended results of the intervention – both positive and negative. Ask experienced people in your field to review risk assessments or findings from similar interventions to help identify these.

What type of evaluation should I conduct?

Evaluations fall into two broad categories:

Formative – evaluations conducted during the development of an intervention. These are useful if you want direction on how to improve the design process.

Summative – evaluations completed once an intervention is in place. These will tell you whether you're achieving your purpose.

You choose which to use based on the purpose of your evaluation. Table 1 details the different types of formative and summative evaluations.

When you conduct an evaluation, collecting data from all the people or sources is rarely possible. Instead, you might select a smaller sample that represents the system as a whole.

What do I measure, and how?

What you need to measure, and how, will be informed by the type of evaluation you conduct and your context. The chapters below provide specific guidelines for each type of evaluation. However, in general, you need to consider:

- your sampling strategy – when and how often you will collect your data, from where, and from whom

- the type of measures, indicators or metrics you will use

- the data collection method – how you will collect or retrieve the data about activities, results, context and other factors

- data analyses and reporting methods – the best way to synthesise and present your findings.

When you conduct an evaluation, collecting data from all the people or sources is rarely possible. Instead, you might select a smaller sample that represents the system as a whole. There are two types of sampling methods:

- **Probability sampling** involves random selection and allows you to make strong statistical inferences about the whole system.

- **Non-probability sampling** involves non-random selection based on convenience or other criteria, allowing you to collect data easily. However any inferences you make are weaker than with probability and your conclusions may be more limited.

How you choose to monitor your evaluation elements will depend on your context – what you are evaluating, the type of evaluation and the accuracy required.

- **Measures** use standard units to express the size, amount or degree of something.

- **Indicators** show the state or level of something.

- **Metrics** are a system or standard of measurement commonly used for assessing, comparing and tracking performance or production.

Although measures tend to be the most precise, their collection may be the most complex and time consuming. They may not be the most relevant or useful for your context.

A simple example – cat outdoor pen – illustrates this difference. To construct one, a builder needs to know the dimensions, i.e. length, width and height in centimetres (measure). However, a cat owner may only want to know how many cats can be accommodated (indicator). To compare the construction safety of various designs, an owner may consider the safety standards for each and/or the material used (metric).

You may need to collect information from individuals or groups (interviewing, focus groups, questionnaires), make observations (directly, through using time-lapse photography to discern changes over time) or examine existing documents and data (literature reviews, project records, big data).

For more specific information, see chapters below.

three

Needs assessment

A 'needs assessment' determines if an intervention is required, who needs it, how great the need is and what can be done to best meet it.

Why conduct a needs assessment?

A needs assessment is typically conducted at the beginning of a project to plan new work. You collect information to:

- define the nature and extent of the problem
- identify priorities and objectives for an intervention
- identify the target individuals / population.

It can also provide baseline data to use in other forms of evaluation or to better understand the context for continuing the intervention.

Determining a 'need' can be a subjective process, so a robust needs assessment should include information from multiple sources, including:

- knowledge and skills from experts in the field
- current best evidence from research in the field
- experience from stakeholders and the target audience.

A needs assessment leads to a more effective, transparent and defensible use of resources. Interventions informed by a needs assessment will be better targeted and planned (rather than reactive) and aligned with community needs and agency priorities.

How do I conduct a needs assessment?

It is best to conduct a needs assessment as a systematic process (Rossi, Lipsey, & Freeman, 2004):

1. Decide on its scope – Clarify the purpose of the needs assessment, determine the approach you'll take and decide who will make the assessment decisions.

2. Determine the assessment criteria – i.e. those agreed factors for judging and prioritising the needs.

3. Make a plan to collect the required information – Identify the type of data required, where this can be sourced and how to collect it.

4. Collect and analyse the data.

5. Use results – Determine priorities about resource allocation and potential intervention approaches.

Refer to McLeod, Driver & Hine (2022) for more detail about planning the best approach and data collection techniques.

Case Study 1: Needs assessment

Emily and Lachlan work in a local government agency responsible for enforcing the state's Companion Animal Act. They are aware this Act has been recently amended and are determining if their agency should change or update domestic cat management procedures. They both recently attended a workshop where they mixed with employees from similar agencies and learnt about a wide range of cat management approaches in other areas.

Their own agency's records indicate that community complaints about nuisance cats have increased over the past two years. Emily recently attended a local conservation group's meeting demanding action about managing a local nature reserve – home to several vulnerable wildlife species. Just last week a local vet mentioned, in an informal pub conversation, that three litters of unidentified kittens had been left at their practice.

To formalise and make the collection of this ad hoc information more transparent, Emily and Lachlan decide they should conduct a systematic 'needs assessment'. In this way, their agency's policy, strategy and interventions will be better targeted, planned (rather than reactive) and aligned with community needs, as well as agency priorities and resources.

Step 1 (Scope): The purpose of the needs assessment is to define the nature and extent of the domestic cat problem in their local community and identify priorities to inform their agency's cat management programs, policies and services. They want to better understand the context of what is currently being done and plan for changes if required.

Emily and Lachlan acknowledge that determining the nature of the problem and identifying the priorities (i.e. the 'needs') will be a subjective process, as different people are likely to have varied views on what the community requires and on domestic cat wellbeing. To ensure the full range of issues is identified, they plan to gather evidence from many stakeholders, including their agency, cat management experts, other agencies involved in cat management, cat advocacy groups, veterinarians, cat owners and the general public.

Step 2 (Assessment criteria): Emily and Lachlan create a list of factors to help them prioritise their needs (Smart, 2019). They consider:

- the problem itself – who and what is being affected (number and frequency), the impact being felt, the complexity of the problem and how easy it is to change

- the context and community – if there is public or political support, whether other groups are addressing the need (opportunities for collaboration and resource sharing), how the problems align with existing priorities and legal requirements

- their agency and resources – the skills and expertise of agency staff, senior management support, availability of resources and current presence of effective interventions.

Step 3 (Plan to collect information): Emily and Lachlan decide on a mixed methods approach. They will gather information from agency records, conduct face-to-face interviews with key stakeholders and survey community members via an online questionnaire.

Step 4 (Collection and analysis of data): Once Emily and Lachlan develop their plan, they commence data collection. They have to be flexible with some aspects. As face-to-face interviews breach their agency's COVID-19 protocols, they organise online interviews.

During the interviews, they become aware of a new aspect of the problem. There are several cat colonies in their area being cared for by dedicated carers. These cats have to be included in their data collection too.

The main objective of the data analysis is to generate a list of the different needs and/or identified assets. As it is quite extensive, Emily and Lachlan consolidate the needs into similar groups so they have a manageable amount of data to present. For example, they identify:

- mandatory needs for responsible cat ownership, as dictated by the current legislation (e.g. microchipping, desexing and registration of pet cats)

- optional aspects (e.g. cat containment, declaration of cat-free areas)

- needs for managing nuisance cats (e.g. reporting, trapping, housing) and stray (un-owned cats).

Their formal report presents findings to decision-makers within their agency. A shorter, more accessible summary is provided for stakeholders and interested community members.

Step 5 (Use of results): Emily and Lachlan, along with the decision-makers, can now use the criteria developed in Step 2 to prioritise identified needs for action. The results will inform the agency's policy planning and cat management strategies.

four

Process evaluation

Process evaluation examines how an intervention was conceptualised and implemented (i.e. the processes for developing and executing the intervention). It can be done either as a one-time assessment or continuously throughout the process.

Why conduct a process evaluation?

An intervention based on weak or faulty conceptualisation and/or implementation has little prospect of achieving the intended results. Assessing these processes allows you to identify weaknesses and faults. You can refine what you're doing and improve any future processes.

Earlier, we discussed how important it to understand the theory or logic behind your intervention (i.e. conceptualise a logic model). This helps you identify important evaluation questions and is essential in your intervention attaining the desired results. You need to assess how good it is – in particular, how well it is formulated, whether it is plausible and feasible for improving the targeted behaviours, and whether it is valuable.

However, an intervention needs more than a good plan of attack to achieve the desired outcomes. Execution is key. Implementation may seem straightforward, but you may have to contend with unforeseen circumstances that compromise delivery and performance. Assessing and identifying these issues allows you to improve the delivery process, increasing the likelihood that your intervention will achieve the desired outcomes.

How do I conduct a process evaluation?

Assessing the development of your intervention

Before you can assess your intervention's theory, you must be able to express it clearly and completely enough to stand for a review. Three components to include in this description are (Rossi, Lipsey, & Freeman, 2004):

1. **Impact theory** – a description of cause-and-effect sequences, where intervention activities are the instigating causes and the changed behaviours or benefits are the effects produced

2. **Service plan** – a description of the sequence of events used to engage the target population with the intervention

3. **Organisational plan** – a description of how the functions and activities of the intervention should perform and the human, financial and physical resources required.

There are several approaches to assessing your intervention's theory:

* comparing the type of intervention with the needs you are addressing, using the results from a completed needs assessment

* comparing the type of intervention with the expected behaviour changes, using evidence from behaviour change research or documented practice elsewhere

* appraisal by stakeholders and others of the clarity, plausibility, feasibility and appropriateness of the intervention.

Supplement any of these with direct observation, if it is possible.

Assessing the execution of your intervention

Before you can evaluate your execution, you must determine the assessment criteria. These may include stipulations from the program theory, administrative standards, applicable legal, ethical or professional standards and after-the-fact judgement calls (Rossi, Lipsey, & Freeman, 2004).

People commonly use:

* **Implementation evaluation** – assesses whether the intervention is delivered as intended to the target audience. This may be a stand-alone evaluation that asks questions about the operations and delivery of the intervention, or be done in conjunction with an impact evaluation to determine how the delivery of the intervention complements findings about its impact on the target audience.

* **Process monitoring** – focuses on assessing the service and organisational plans and asks questions about coverage, bias, resource use, delivery and support functions. It also identifies any shortcomings that prevented the intervention from being delivered. Process monitoring can serve different purposes – evaluation, accountability and intervention management. The data required for each is fairly similar and can be integrated into any routine management information collection and reporting system.

Main data sources include target audience surveys and community and project records. Refer to McLeod, Driver & Hine (2022) for more detail about survey design.

Analyses typically address descriptions of how the intervention was implemented, comparison across sites, and how delivery conformed to design. For these and other analysis tips, refer to Rossi, Lipsey & Freeman (2004).

Case Study 2: Assessing intervention theory

Emily and Lachlan's needs assessment (see Case Study Box 1: Needs assessment) identified cat containment as a priority. They decide to explore the suitability of their agency's current intervention (a 'fact sheet' with educational information and persuasive messaging on their website). Is this the best way to encourage cat owners to adopt full-time containment methods (i.e. is the intervention theory sound?) or will they need to improve the design?

Emily and Lachlan have no outcome or impact evaluation feedback available for their current intervention so, as a first step, they decide to consider evidence from best-practice principles in behaviour change research and other documented approaches. To compare different interventions with expected behaviour changes, they choose the Behaviour Change Wheel framework and its associated Capability-Opportunity-Motivation-Behaviour model (Michie, Atkins & West, 2014). This is helpful in both designing interventions and evaluating whether a particular intervention employs appropriate behaviour change techniques (BCTs) to achieve required outcomes.

Step 1 (Describe their intervention theory): Emily and Lachlan create a simple logic model to clearly express the theory of their intervention (Figure 2).

Step 2 (Characterise their intervention): Interventions can be characterised, and their content described, by identifying the type of BCTs employed in their design. Emily and Lachlan identify five BCTs using Hine, McLeod and Driver (2022) as a guide:

1. Educate (information on why cat owners should contain their cats)
2. Train (instruction on how to contain cats)
3. Inform (consequences of not containing them)
4. Discuss (comparison information of other cat owners, debunking misinformation)
5. Persuade (message framing – wildlife loss, welfare benefits, story-telling).

Step 3 (Investigate the functions played by BCTs): Specific BCTs work best with factors that influence the adoption of the desired behaviour. For example, education works best if a person is unaware of a problem but won't help if they can't afford to behave in the desired manner. Again, using Hine, McLeod and Driver (2022) as a guide, Emily and Lachlan identify that the educational and training component of their intervention could address cat owners' capabilities to contain their cat. Informing, discussing and persuading may assist their motivation.
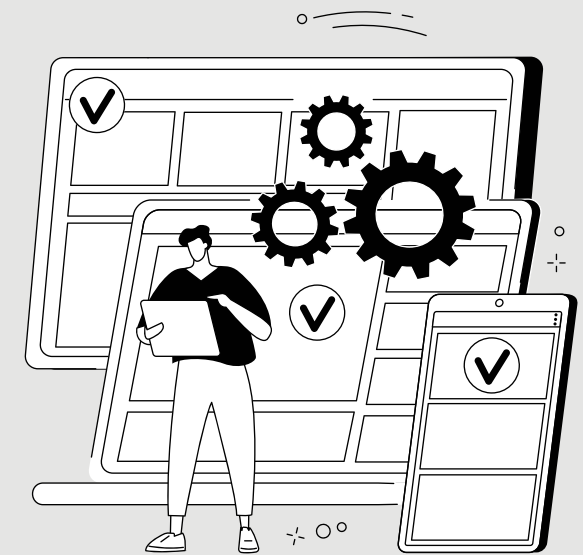
Figure 2 Simple logic model of cat containment intervention

Inputs	Outputs		Outcomes		
Investments	Activities	Participants	Short term	Medium term	Long term
Agency provides cat project officer, fact sheet designer, printing and website access	Project officer promotes fact sheet at community events, hard copies distributed to local vet clinics, animal shelters	Cat owners	Creating awareness, knowledge and skills about cat containment	Changes in cat containment behaviour	Cultural shift in attitudes to cat containment, decrease in wildlife impacts, increase in cat welfare outcomes

Step 4 (Compare to expected behaviour changes): Emily and Lachlan consult research literature to discover the types of barriers and drivers expected for cat containment (e.g. McLeod, Hine & Bengsen, 2015; Elliott, Howard, McLeod & Bennett., 2019). They find a range of factors, including capabilities (e.g. awareness, confidence), opportunities (cost, inability due to renting) and motivations (beliefs about cat well-being, it not being mandatory).

Step 5 (Analyse results): Emily and Lachlan conclude that their current intervention inputs and outputs may fall short of achieving their intended outcomes. For example, it does not address all expected factors that could influence behaviour change. They'll need to better understand barriers and drivers facing cat owners in their community and maybe add some other BCTs like commitments, incentives and support.

Case Study 3: Assessing implementation

Working with the local welfare animal shelter, Emily and Lachlan develop a new intervention for encouraging full containment. It incorporates both educational information, instructions on how to best keep a new cat contained and a support program to address the personal circumstances and needs of new cat adopters.

Step 1 (Scope): As the support aspect of this intervention is crucial to their overall evaluation, Emily and Lachlan want to assess whether it is being delivered as intended.

Step 2 (Determine assessment criteria): The assessment criteria Emily and Lachlan use are guided by the support protocol developed for this intervention, administrative standards of the shelter, and human research ethical standards.

Step 3 (Plan to collect information): Emily and Lachlan decide on a two-stage plan. Within a month of implementation, they conduct a stand-alone evaluation using direct observation – focusing on the practical operations and delivery of the intervention.

Shelter staff are to routinely record information after each encounter. This will provide data about resource use, delivery and support functions, coverage and bias of the program, and identifies any shortcomings that prevent the intervention from being delivered as intended.
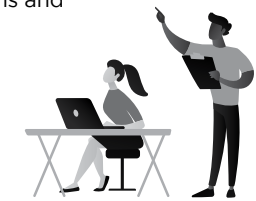
Step 4 (Collection and analysis of data): Information collected from observations of the support protocol highlights some unforeseen issues. Despite training courses for shelter workers, several volunteers are still not confident in delivering the intervention. Also, Emily and Lachlan observe that the routine reporting system cannot capture all required information. Consequently, they refine the process to make it easier for shelter staff and volunteers to deliver the intervention and report.

five

Outcome evaluation

This type of evaluation investigates how well an intervention has achieved its outcomes, i.e. an observed characteristic or behaviour of the target audience.

The outcomes of your intervention usually lie in its impact theory (i.e. in the description of the intervention's cause-and-effect sequences, where the intervention's activities are the instigating causes, and the benefits or changed behaviours are the effects).

Outcomes are not to be confused with 'outputs', which are the services delivered by an intervention. Outcomes relate to the benefits of these outputs, not just their provision. *For example, you may hold a workshop attended by 50 target audience members. This workshop is part of the intervention process, i.e. part of the service delivery, an output. In contrast, the training benefit – improved capability (skill and confidence) of the participants – is an outcome.*

Why conduct an outcome evaluation?

The ultimate goal of your intervention is to bring about change – to affect some problem beneficially. To determine success, you need to measure the outcomes, the state of the target audience or the problem that the intervention intended to change. Outcome monitoring can also assist in refining and improving interventions.

Definitions (Rossi, Lipsey & Freeman, 2004)

- **Outcome level** – the status of an outcome at some point in time (*e.g. the number of the target audience performing the desired behaviour*).

- **Outcome change** – the difference between outcome levels at different points in time (*e.g. the difference between the number of the target audience performing the desired behaviour before the intervention commences and the number after six months of delivery*).

- **Intervention effect** – the portion of an outcome change specifically attributed to the intervention – not the influence of some other factors.

How do I conduct an outcome evaluation?

To conduct an effective outcome evaluation, consider (Rossi, Lipsey & Freeman, 2004):

1. **Relevant outcomes of the intervention to be monitored** – These can be identified by consulting the project's impact theory and other documents, engaging with stakeholders and reviewing pertinent research.

2. **Choice of measures and indicators (outcome variables)** – These need to be reliable, valid and sufficiently sensitive to detect the expected level of outcome change. It is often advisable to use multiple variables to correct for possible weaknesses in one or more of your chosen variables and improve the strength of your findings.

3. **How these outcome variables are collected** – This will depend on the type and complexity of the selected variable. You can collect these throughout the intervention, i.e. outcome monitoring (e.g. monthly collection of participation rates, or daily hits on a website) or at completion.

4. **How the results are interpreted** – Outcome evaluations report on the extent of the change in outcome levels at different points in time. Expert judgement and benchmarking (comparing outcome levels and changes to similar programs) are helpful techniques.

You cannot use these to determine the portion of an outcome change attributable directly to the intervention (causal intervention effect). Many factors in the environment can affect outcomes. Estimation of the intervention effect (commonly referred to as 'impact evaluation') is more demanding (see next chapter).

Case Study 4: : Outcome evaluation

In Case Study Box 3, Emily and Lachlan evaluated the implementation of their new intervention for full cat containment. However, they also needed to evaluate the outcomes of this intervention, so here they develop a plan, guided by the study by McLeod et al. (2020).

Step 1 (Outcomes to be measured): The outcomes to be monitored were the intentions of participants to keep their cat contained at the time of adoption, the changes in knowledge, skills and confidence of participants to contain their cat, and their cat containment behaviour eight weeks later.

Step 2 (Measures/indicators): Emily and Lachlan recorded the participants' intentions and changes in knowledge, skills and confidence using a 5-point Likert scale. Cat containment behaviour was determined from a series of indicator questions tested in multiple prior studies, relating to how and where the cat spent their typical day.

Step 3 (Collection of data): All data was self-reported, using a face-to-face interview at the time of adoption and an online questionnaire eight weeks later.

Step 4 (Interpretation of results): To see if the intervention was having an influence on cat containment (their outcome), Emily and Lachlan investigated the extent of the change in participants' containment knowledge, skills and confidence, along with containment intention and behaviour over the eight-week period of the evaluation. They found varying improvements in knowledge, skills and confidence among individual participants from the intervention, however other factors not targeted by the intervention (e.g. managing cat behaviour) prevented some participants from containing their cat once they got it home. These results led them to consider other additional interventions that would be required to help cat owners to contain their pets.

six

Impact evaluation

Impact evaluations aim to determine what effects an intervention has on its intended outcomes, and any unintended side effects, i.e. to show, with a high degree of certainty, that the intervention caused a particular effect.

Why conduct an impact evaluation?

When assessing how well an intervention worked you need to establish, with a high degree of certainty, that it specifically caused the change in the measured outcomes (e.g. altered the target audience's behaviour or outcomes in the target problem) – not some other independent factors.

Impact evaluations can occur at many points in the course of an intervention's implementation:

- You can pilot a new intervention on a small scale to demonstrate if it has the intended effects. Refinement after a smaller pilot study is easier and cheaper than for a larger project.

- Once the intervention rolls out on a broader scale, you may evaluate its impact to ensure there are no unforeseen problems on this bigger scale.

- Ongoing interventions also benefit from periodic impact evaluations to demonstrate their continuing effectiveness and relevance and identify any new problems that arise over time. This provides the means to defend the intervention against alternate suggestions or funding changes.

How do I conduct an impact evaluation?

The more rigorous the research design of your impact assessment, the more confident you can be about the validity of your estimate of the intervention's effects.

When designing an impact evaluation, you are often faced with two competing pressures:

- Evaluations need to be conducted with sufficient rigour so relatively firm conclusions can be reached.

- Practical considerations of time, money, cooperation and ethical concerns may limit design options and methodological procedures that can be employed.

Aim for the most rigorous design you can afford.

Control group

Always include a control group – a comparison group that does not experience the intervention. Without a control it is impossible to know whether a change in behaviour or other outcome is due to the intervention or one or more of an infinite number of other factors you cannot contain.

Randomised field experiments

Across many scientific disciplines, randomised controlled trials represent the gold standard for evaluating a treatment or project. The same principles apply to impact evaluations of your intervention. Random assignment of participants to experimental conditions ensures these groups are as similar as possible before delivering the intervention. This means any observed differences between the treatment and control groups should be attributable to the intervention and not pre-existing group differences or other miscellaneous factors.

Alternate designs

When random assignment is not possible, you can use quasi-experimental designs. Quasi-experiments compare naturally occurring or self-selected groups. For example, you could launch an intervention promoting the reporting of stray or feral cats in a community and compare the reporting practices of those who are aware of the intervention and those who are not. A quasi-experimental design will not give you the same level of confidence that your intervention was the main factor driving the outcome change as a randomised control experiment but, in most cases, imperfect evidence is better than no evidence.

Analysis

Use statistical tests to evaluate effects – these help you decide if the measured differences between treatments are 'real' or simply due to chance. In some situations, you can be confident your intervention has been effective. *For example, if you find participation has increased by 80% in your treatment group compared to 10% in the control, you can be reasonably certain you're onto something. However, if you find it only goes up by 10% in your treatment group compared to 5% in the control group, how certain can you be of a meaningful result?* If you are unfamiliar with the many statistical tests and software options available, consult a qualified statistician within your organisation or local university.

Many excellent references can assist you in design and analysis for a wide range of contexts. As a starting point, we suggest Shadish, Cook & Campbell (2002), Rossi, Lipsey & Freeman (2004) or Murnane & Willet (2010).

Case Study 5: Impact evaluation

Emily and Lachlan wish to assess the impact of the persuasive messaging they developed for their new cat containment communications in a pilot study, before committing more resources to implementation. They want to see if messaging affects outcomes in the short and medium terms and also learn, with a high degree of certainty, whether the messaging itself, not some other factor, is the cause. The design of their impact evaluation follows a similar study conducted by McLeod, Hine, Bengsen & Driver (2017):

Step 1 (Research design): Emily and Lachlan randomly assigned their participants to either a treatment group (their persuasive message) or a control group (neutral message).They planned touse mediation and moderation statistical tests to analyse their data.

Step 2 (Measures/indicators): Participants responded to standard questions gauging their current cat containment perceptions, intentions and behaviour.

Once presented with their particular randomly assigned message, they answered standard questions about how that message influenced motivation to contain their cat, and their perception that they could effectively do so (response efficacy) by rating their agreement on a 5-point Likert scale).

Participants again responded to questions about their cat containment perceptions and future intentions.

A follow-up survey after four weeks assessed any further change in intentions and cat containment behaviour.

Step 3 (Collection of data): All data was self-reported using an online questionnaire. An online research panel recruited participants.

Step 4 (Interpretation of results): Emily and Lachlan found that their message increased participants' motivation and response efficacy compared to the neutral message. This improved motivation and response efficacy predicted an increase in cat containment behaviour.

From these results, Emily and Lachlan were confident to go ahead and launch their message ona broader scale. They planned to evaluate the message's impact again at this larger scale, and also evaluate longer-term outcomes (e.g. decrease in wildlife impacts, increased cat welfare).

seven

Efficiency analysis

Efficiency analysis relates the cost of an intervention to its beneficial outcomes. Examples include:

- **cost-benefit analysis** – compares benefits and costs in monetary terms
- **cost-effectiveness analysis** – relates costs (expressed in monetary terms) to pre-defined 'units' of outcome.

Why conduct an efficiency analysis?

Resources are always limited, so it is important to measure the successful outcomes and impacts of interventions and calculate overall benefit per dollar spent. Efficiency analysis can help you defend the implementation of an intervention or choose between competing ones. If two implementation strategies are equally effective, but one costs substantially less, it makes sense to select the cheaper option.

Efficiency analysis can also be conducted before an intervention, i.e. during the planning and design phase. It may help you decide whether to pursue an intervention, especially if it requires extensive commitments or will be difficult to abandon once in place. This analysis relies on the assumption that a positive impact will be achieved, and that anticipated costs and outcomes can be reasonably estimated.

How do I conduct an efficiency analysis?

There are five basic steps to an efficiency analysis:

1. Outline your objectives and what parameters you will need to measure.

2. Identify costs and benefits.

3. Calculate costs and benefits, i.e. assign a value, either monetary or another 'unit'.

4. Tally the total values and compare.

5. Analyse the results and make an informed, final recommendation.

When identifying costs/benefits, consider:

- **direct costs/benefits** – related to the development and implementation of the intervention (e.g. labour and materials costs or increased revenue from new product sales)

- **indirect costs/benefits** – usually fixed in nature, and may come from overheads of the organisation involved with implementation (e.g. utilities or rent)

- **intangible costs/benefits** – those costs that may be difficult to measure and quantify (e.g. shifts in target audience satisfaction or employee morale)

- **opportunity costs/benefits** – lost benefits or opportunities that arise when one intervention is pursued over another (e.g. loss of a direct subsidy that no longer applies).

The accuracy of efficiency analysis depends on assumptions made and techniques used, i.e. how costs/benefits are estimated, how secondary and distributional effects are handled, and how future costs and benefits are discounted to reflect present values. Chapter 11 in Rossi, Lipsey & Freeman (2004) offers an excellent starting point. Guidance from an expert in this area is also important.

Assigning monetary values to non-tangible concepts can be complex and controversial. Cost-effectiveness analysis is feasible when benefits cannot be calibrated in monetary terms. It allows you to compare interventions with similar goals and outcomes in terms of relative efficiency. It can also be used to analyse the relative efficiency of variations to an intervention.

For example, if intervention A promotes microchipping among cat owners via one-on-one outreach and intervention B uses messages on social media, you can compare the relative cost for each extra cat that is brought into the clinic for microchipping. You will not need to assign a benefit monetary value for each microchipped cat.

> *Cost-effectiveness analysis is feasible when benefits cannot be calibrated in monetary terms. It allows you to compare interventions with similar goals and outcomes in terms of relative efficiency.*

Case Study 6: Cost-effectiveness analysis

Another priority identified in Emily and Lachlan's needs assessment (see Case Study Box 1: Needs assessment) was cat desexing and preventing unwanted animals. Talking with colleagues from other agencies, they become aware of the National Desexing Network (NDN) which offers an alternative desexing intervention to their agency's 'collect/hold/euthanase' model (National Desexing Network, 2021). They conduct a cost-effectiveness analysis to choose whether the NDN model might be a better approach.

Step 1 (Outline objectives): This evaluation aims to compare their agency's cost for each desexed cat across two alternate models (NDN vs current 'collect/hold/euthanase').

Step 2 (Identify relative costs and outcomes): The relative outcome they will use to compare the programs is the (dollar) cost to their agency per desexed cat. The costs of the NDN program will include the direct cost of funding their share of the subsidy, staff time in negotiating partnerships with local vet practices and other local organisations, and indirect costs like administration. Their current model includes both direct and indirect costs associated with collecting unwanted animals from owners, running a holding facility and euthanising unwanted animals. An intangible cost to consider is the negative emotional impact on staff at the holding facility.

Step 3 (Calculate costs and compare values): Emily and Lachlan calculate, tally and compare their values for each program (Table 2).

They calculate that their agency could save between $160 and $360 for every unwanted kitten. Annual savings, based on 300 cats being desexed on the NBN program (at the cost of $18,000) would be conservatively in the range of $48,000–$108,000 (assuming each cat would produce one litter of unwanted kittens).

Table 2 Costs per cat for each program (after National Desexing Network, 2021)

Desexing program	Cost per cat
NDN program	• $60
Collect/hold/euthanase	• $220 (held for 4 days) • $420 (held for 14 days)

Step 4 (Make recommendations): From these results, Emily and Lachlan recommend their agency further explore participation in the NDN program.

eight

References

Alter, T., Driver, A., Frumento, P., Howard, T., Shuffstall, B., Thompson, L., & Whitmer, W. (2017). *Community engagement for collective action: A handbook for practitioners.* Canberra, Australia: Invasive animals CRC. Available at: https://community.invasives.com.au

BetterEvaluation (2014). *Rainbow Framework.* Retrieved February 2021 from https://www.betterevaluation.org/en/rainbow_framework

Elliott, A., Howell, J. T., McLeod, M. E., & Bennett, C. P. (2019). Perceptions of responsible cat ownership behaviors among a convenience sample of Australians. *Animals*, 9, 703-713.

Hine, D. W., McLeod, L. J., & Driver, A. B. (2022). *Designing behaviour change interventions for sustainable land management: A practical guide.* Canberra, Australia: Centre for Invasive Species Solutions. Available at: https://community.invasives.com.au

McLeod, L. J., Evans, D., Jones, B., Paterson, M., & Zito, S. (2020). Understanding the relationship between intention and cat containment behaviour: A case study of kitten and cat adopters from RSPCA Queensland. *Animals*, 10, 1214-1230.

McLeod, L. J., Hine, D. W., & Bengsen, A. J. (2015). Born to roam? Surveying cat owners in Tasmania, Australia, to identify the drivers and barriers to cat containment. *Preventive Veterinary Medicine*, 122, 339-344.

McLeod, L. J., Hine, D. W., Bengsen, A. J., & Driver, A. B. (2017). Assessing the impact of different persuasive messages on the intentions and behaviour of cat owners: A randomised control trial. *Preventive Veterinary Medicine*, 146, 136-142.

McLeod, L. J., Driver, A. B., & Hine, D. W. (2022). *Designing community surveys for behaviour change research: A practical guide. Canberra, Australia: Centre for Invasive Species Solutions.* Available at: https://community.invasives.com.au

Michie, S., Atkins, L., & West, R. (2014). *The behaviour change wheel. A guide to designing interventions.* UK: Silverback Publishing.

Murnane, R. J., & Willett, J. B. (2010). *Methods matter: Improving causal inference in educational and social science research.* New York: Oxford University Press.

National Desexing Network (2021). *The benefits of councils funding NDN co-operative desexing programs.* Helensvale, Queensland: National Desexing Network. Available at https://www.g2z.org.au/

Rossi R H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach.* Thousand Oaks CA.: Sage Publications.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-experimental Designs for Generalized Causal Inference.* Boston: Houghton Mifflin.

Smart, J. (2019). *Needs assessment. Practice resource.* Southbank, Victoria: Australian Institute of Family Studies.

"Evaluating Behaviour Change Interventions:
A Practical Guide" introduces key principles for
developing evaluation plans. Importantly, the
guide offers practical advice for designing and
implementing evaluation plans and analysing the
data they generate.

This book is intended for natural resource management organisations and practitioners that
work with the general public and interested stakeholders on sustainable land management.