Evaluation design

Key information in this section

- Design criteria for evaluations

- Information requirements including using multiple lines and levels of evidence

- Identifying performance measures

- Using an evaluation panel

- Developing an evaluation plan

Evaluation design is concerned with the detailed planning of the evaluation. It builds on the evaluation context to specifically identify practices, processes, timing and responsibilities for completing an evaluation. The detailed information gathered at this stage can be compiled into an evaluation plan for the particular type of evaluation and contribute to the portfolio of evaluations discussed below.

No single evaluation design applies across all evaluations. The design will depend on the evaluation context documented by the evaluation team, time constraints, existing information and the resources available.

Evaluation design criteria

Information collated when determining the scope of an evaluation can be used to help define evaluation design issues. All design options have strengths and weaknesses when considered in the context of specific evaluation requirements. Criteria which could be relevant when considering the design of an evaluation are:

- relevance to the evaluation purpose

- cost-effectiveness

- validity of the information

- ethical issues around how the information is to be used.

The evaluation team should be conscious that, as more information becomes available or gaps are identified, the original design may need to be adjusted.

Evaluation design issues relating to each of the three types of evaluation are discussed in:

- Design for appropriateness evaluations
  (www.environment.nsw.gov.au/resources/4cmas/08640appevaldesign.pdf)

- Design for efficiency evaluations
  (www.environment.nsw.gov.au/resources/4cmas/08641efficevaldesign.pdf)

- Design for effectiveness evaluations
  (www.environment.nsw.gov.au/resources/4cmas/08655effectevaldesign.pdf).

Information requirements

Evaluations will use both existing and new information in their assessment of a project or program.

The earlier work in defining the evaluation context will determine the evaluation type (or combination of types) required. From this information and using the above design criteria, an evaluation team can start to define the methods needed.

It is most likely that a combination of methods will mean a mix of qualitative and quantitative information will be analysed. This is often referred to as a 'mixed method' approach to evaluation and a combination of both sets of information may lead to a richer base of evaluative information.

Types of information in a mixed method approach

Descriptions illustrate or describe what is happening in the program. Examples include:

- the process used to deliver an incentives program
- the outputs from the program
- what may happen to incentive recipients after they complete their obligations under the program.

Judgements are made when a program's performance is being compared with standards, targets or benchmarks. Examples include:

- an NRC audit report against the State-wide Standard for NRM
- progress towards catchment targets.

Interpretations are based on inferred relationships and used to explain whether the measured outcomes of a program can be attributed to the program.

An evaluation may draw on a mix of these different information types. A simple example could be descriptions of outputs from a riparian incentive program. Descriptions of downstream sediment loads would be combined with knowledge from previous riparian management studies and literature to interpret the relationship between the incentive program and water quality outcomes measured downstream.

While it is not important to specifically classify information into the types described above, it is important to understand the roles different information types play and how most evaluations will draw on multiple lines of information from each.

Using multiple lines and levels of evidence

With an understanding of the types of evaluation designs and the diversity of activity and complexities associated with a CAP, it is obvious that no single source of evidence or analytical design will be able to address all evaluation issues. Multiple lines and levels of evidence (MLLE) is the evaluation method recommended for CAP evaluations because it can infer relationships using various sources of information, existing or new, based on an assessment against criteria.

MLLE was first proposed by Hill (1965) in the medical field and has since been used in human and ecological risk assessments (Culp et al 2000; Fairbrother 2003). It is now being adapted for NRM (Adams 2003; Young et al 2006).

The MLLE framework uses information from a variety of sources. This is first assessed against criteria and then combined to address evaluation questions. Table 1 defines the lines and levels of evidence and provides examples of each.

Multiple sources of evidence can be used to address an evaluation question. Evidence sources are generated from existing data or information or new data or information created for the

evaluation. From a cost-effective perspective, existing sources should always be reviewed first. If the weight of evidence is insufficient, new data sources should be sought.

Table 1: MLLE framework components and examples relevant to a CAP

Framework component	Description	Example
Multiple lines of evidence	Evaluation evidence from various sources. Can be a mix of qualitative and quantitative, existing or specifically generated.	Evidence that riparian management improves bank stability from: Abernethy, B & Rutherfurd, I 1999, 'Riverbank reinforcement by riparian roots' in I Rutherfurd & R Bartley (eds) *Proceedings of Second Australian Stream Management Conference*, Adelaide, vol. 1, pp.1–8.
Multiple levels of evidence	Criteria used to determine confidence or strength of inference that the information can address the evaluation question. Sometimes referred to as 'weight of evidence'.	Is the response of riparian management specific to CMA catchments? Is the response to good riparian management consistent?

Examples of existing data sources are detailed in Table 2 while Table 3 lists examples of new data sources.

Table 2: Examples of existing data sources

Potential existing sources	Description
Catchment target performance measures (PMs)	The resource condition performance measures are considered 'existing sources' as they are required by the intergovernmental agreement and monitoring and reporting systems should be in place. In addition, some catchment PMs may use data from programs with state-wide targets. At a catchment scale, these state monitoring programs may need to be supplemented with additional data. PMs compared with targets will provide a description of the condition of the resource and a judgement about how it is performing against the catchment target. This information alone will not allow for cause-and-effect inference between investment and catchment target.
Published literature	Common source of evidence Review that identifies scientific evidence for a link between an issue or action and its response Review needs to consider the applicability of the findings to the CMA's landscape and natural resource management issues.
Existing catchment reports	As with published literature, but more relevant to the catchment concerned

Potential existing sources	Description
CMA annual reports and annual implementation plans, CAP and investment strategy	Investment trends provide information on outputs achieved for the investment and links to catchment and management targets. Analysis of this information in a program logic framework particularly provides information on the efficiency and effectiveness of a CAP.
Scenario modelling (if already exists)	The use of existing well-calibrated scenario models can predict resource condition that results from the implementation of management actions.
CAP project reports or LandCare group diaries, etc.	The management of this quantitative and qualitative information so that is easily accessed and reviewed could also assist in evaluations against the management outputs and management targets. At a CAP scale it will be important to review the results hierarchy and evaluation questions to determine which projects to focus on. Information from projects may also supplement case studies. The Hunter–Central Rivers CMA publication *Keeping a project journal* (Wark 2005) is aimed at LandCare group projects. Review of this information may be used with case studies or information on barriers or relevant monitoring.

Table 3: Examples of new data sources

Potential new sources	Description
Case studies	Case studies provide an in-depth understanding of a single issue or case and its likely relationships to actions. Case studies use a variety of methods to collect data, focused on the particular questions being investigated by the case. May range in complexity and be unnecessarily resource-intensive if not well designed.
Scenario modelling (if new work)	Resources are required to develop and validate a model which can predict the resource condition that results from the implementation of management actions.
Surveys or interviews	Surveys can range from highly structured questionnaires to non-directive interviews. Can provide qualitative and quantitative information from stakeholders depending how they are designed.
Externalities	Activities that are outside the control of the CMA but may impact (positively or negatively) on the ability to achieve the management and catchment targets. Externalities are identified within the program logic tables. Important to identify early, but only monitor if other levels of evidence indicates the need.

The type of evaluation will determine the lines of evidence (including constraints) that will need to be sourced for the evaluation. Some of the analytical tools or methods that may be used are then applied before bringing the different lines of evidence to the table for assessment against the evaluation questions posed. Some of these evaluation analysis tools available are detailed in the [evaluation tools](www.environment.nsw.gov.au/4cmas/evaltools.htm) (www.environment.nsw.gov.au/4cmas/ evaltools.htm). This list is not intended to be definitive but provide a starting point to the range of tools available.

Using MLLE early to determine monitoring needs

It is important that the strength of inference from existing information sources is understood as early as possible in the CAP implementation phase. For example, assuming strong lines of evidence from literature and past catchment studies on the benefits of stabilisation and revegetation of highly erodible soils may reduce the need to monitor other data sources. This will free up expenditure which can be used to gather new data on other catchment targets as required.

Initial application of the lines of evidence criteria is recommended as part of a gap analysis when determining monitoring needs. This would be undertaken by:

- determining context and specifically confirming the types of evaluation to be undertaken

- defining key evaluation questions and information needs

- collating existing information

- reviewing existing information against lines of evidence with an evaluation panel to determine confidence in the available information to inform the evaluation questions

- documenting this process

- determining how best to address any information gaps and implementing programs.

Considerations when applying MLLE

Young et al (2006) identified numerous issues to be considered when using the MLLE framework. These have been adapted below for application to the CAP process.

Clearly defined evaluation questions: If the question in the program logic table is too broad, the existing sources of data may not be specific enough to enable rigorous application to the criteria. This may also increase unnecessary investment in new data sources when a more clearly defined question would have avoided extra costs.

Determining which literature or studies are relevant to the questions: This does not mean that only literature that supports a positive answer to the question is sought. It is important to consider both sides. The issue is about being able to use the literature in the context of the specific landscape or application. The use of the literature should be documented, as well as the reasons why it was chosen.

Criteria to inform interpretation: The strength of the evidence provided by collated existing and new data needs to be assessed against each evaluation question. Table 4 presents recommended criteria for assessing relevance of evidence (adapted from Chessman 2005; Downes et al 2002; Young et al 2006).

Table 4: Criteria for assessing relevance of evidence

Criteria	Question posed by criterion	Example of response to criterion from different lines of evidence*
Credible linkage	Is the evidence relevant to the issue, i.e. is this evidence expected in the study area?	Literature (references) demonstrates there is a relationship between improved riparian management (activity) and sediment nutrient fluxes (response) of waterways. Local catchment studies (referenced) support the findings of the above literature. Performance measure monitoring for catchment targets is consistent with investment trends and outputs for riparian management.
Presence of a response	Is there a reliable explanation for the evidence in the study area?	Performance measure monitoring for catchment targets linked to SedNet scenario modelling within a case study provides evidence for the management of sediments and nutrients within the catchments. Performance measure monitoring for catchment targets is consistent with investment trends and outputs for riparian management for those estuaries. Reported catchment investigations for like systems.
Evidence of 'intervention-response' relationship with activity	Is this evidence likely only under different spatial or temporal situations?	Performance measure monitoring for catchment targets is consistent with investment trends and outputs for riparian management. SedNet scenario modelling highlights how the application of relevant management actions in priority catchments will have the greatest reduction in sediment inputs to estuaries. Performance measure monitoring for catchment target linked to SedNet scenario modelling within a case study provides evidence for management of sediments and nutrients within the catchments.
Consistency of association	Does the expected response always occur in the presence of the activity? Is there a credible linkage between intervention and response?	Performance measure monitoring for catchment target linked to SedNet scenario modelling within a case study provides evidence for management of sediments and nutrients within the catchments. Performance measure monitoring for catchment target in particular estuaries is consistent with investment trends and outputs for riparian management.
Consistency	Are there consistent findings across different lines of evidence?	All lines of evidence tested against criteria and show consistent findings.

* This column provides an example only of how existing and new sources of information may be used with the criterion.

Identifying performance measures

Performance information needs have been recorded in the program logic table. The measures used may be broad and combine several indicators. Running monitoring programs across all of these would be costly and so criteria are used to decide which will be the most appropriate to use. This is particularly the case with performance measures against high-level and intermediate outcomes.

When identifying specific performance measures from a large number of possibilities, it is important to select measures that are reliable and will meet the performance measurement needs of the project or program now and into the future. The practicality of using measures should also be considered.

A template for identifying performance measures is available in the evaluation tools (www.environment.nsw.gov.au/4cmas/evaltools.htm).

Using an evaluation panel

Use of an expert evaluation panel is recommended for significant evaluations. It allows a wide range of expertise to be brought together to assess available information and develop findings and recommendations.

The size of the panel and the multi-disciplinary nature of expertise needed will depend on the evaluation being undertaken. For example, a wide range of internal and external expertise would be appropriate for an effectiveness evaluation of CAP targets. A smaller panel with a shorter operational time frame may be appropriate for an annual review of the performance of the investment strategy and may involve a cross-section of people from the investment, operations and planning areas of a CMA.

Benefits of using a representative and well-run evaluation panel include:

- A panel brings together a range of views, experiences and knowledge for making judgements on the evaluation information.

- A panel develops valid and defensible judgements or conclusions.

- It develops or enhances networks and knowledge for the CMA.

- An evaluation panel is a more cost-effective approach than outsourcing the whole evaluation, due to the contribution from the CMA and its local knowledge. It is also more likely that the CMA will have greater ownership of the results and the adaptive management processes.

- The panel process will bring together local and internal knowledge through CMA contributions as well as broader experience and knowledge.

Establishing an evaluation panel (www.environment.nsw.gov.au/4cmas/tipsevalpanel.htm) has advice on setting up and supporting an expert body.

Developing an evaluation plan

An evaluation plan should be developed to outline and map the requirements and implementation of each evaluation process. Whether an evaluation process is being designed to address the needs of a CAP evaluation or those of a small project, the dimensions of the plan will be similar. The size and relevance of the evaluation will influence the detail within the plan.

Portfolio of evaluations

A number of CMAs are considering the model of a high-level evaluation plan or strategy supported by more detailed planning information for evaluations. This approach involves a relatively brief planning document which could refer to the detailed evaluations to be undertaken to support the catchment and management targets and assess major projects.

Example: High-level evaluation plan for Murray CMA

The Murray CMA is developing a high-level evaluation plan or strategy to identify the broad MER requirements, roles and responsibilities for its CAP. It will identify the major types of evaluations required by drivers and stakeholders, their timing and broad requirements.

A brief planning document will refer to the detailed evaluations that have to be undertaken to support the catchment and management targets and assess major projects. The plan will be used as a tool to communicate evaluation issues, inform project management and support any funding bids. It will detail a hierarchy of evaluations which together cover the detailed planning of all required evaluations.

The evaluations that sit under this strategy could be brought together as they are developed to form a 'portfolio' of evaluations. It is important to keep this portfolio manageable to begin with and let it grow as required.

The evaluation plans that sit under the evaluation strategy would then be more flexible documents that represent a lot of the project planning and specific details required to undertake those evaluations.

Once a portfolio of evaluations is established, it will continue to evolve and provide detail of the thinking behind the evaluations. It will also be a resource to inform future development of evaluations and plans.

Elements of an evaluation plan

An evaluation plan should be developed to meet the specific needs identified by each CMA. As such, plans are likely to differ to some degree from each other and also change over time as they evolve. Despite these differences, each evaluation plan should consider the principles and key elements of evaluation.

Part of the planning for an evaluation is the development and documentation of specific roles and responsibilities for players in the process. This includes:

- development of an evaluation team that meets regularly to discuss progress and issues – often an effective way to manage progress and develop skills

- bringing together people relevant to the subject of the evaluation

- negotiation of roles and responsibilities for team members for all elements of the evaluation within resourcing and timing constraints – this can be documented as part of the plan

- development of required communication and awareness strategies

- identification of any training needs for different elements of the evaluation, such as undertaking surveys, doing performance stories

- implementation of a means of reviewing progress against agreed tasks – this will vary in detail depending on the size of the evaluation.

Implementing an evaluation plan

An evaluation plan should be used to guide implementation of evaluation processes. It provides the detail of responsibilities, timing and processes for the evaluation team and so should be carefully managed:

- The plan should be regularly reviewed and adapted where necessary to reflect changes to projects or programs.

- Where changes occur, efficient distribution of the revised plan and communication changes in responsibilities or timing should be highlighted.

- Regular team meetings where issues or problems can be addressed will support effective evaluation processes.

Ensure the timely implementation of evaluation processes as detailed in the plan. In particular, the establishment of relationships to access and share information, identify sources of support and implement monitoring programs are critical to the effective gathering of information.

References

Adams, SM 2003, 'Establishing causality between environmental stressors and effects on aquatic ecosystems', *Human and ecological risk assessments*, vol. 9, no.1, pp. 17–35.

Chessman, B 2005, 'Causal inference in performance evaluation for natural resource management: A discussion paper', unpublished report for Department of Infrastructure, Planning and Natural Resources, Sydney.

Culp, JM, Lowell, RB and Cash, KJ 2000, 'Integrating mecosm experiments with field and laboratory studies to generate weight-of-evidence risk assessments for large rivers', *Environmental toxicology and chemistry*, vol. 19, no. 4(2), pp. 1167–73.

Downes, B, Barmuta, L, Fairweather, P, Faith, D, Keough, M, Lake, S, Mapstone, B and Quinn, G 2002, *Monitoring ecological impacts: Concepts and practice in flowing waters*, Cambridge University Press, Cambridge.

Fairbrother, A 2003, 'Lines of evidence in wildlife risk assessments', *Human and ecological risk assessments*, vol. 9, no. 6, pp. 1475–91.

Hill, AB 1965, 'The environment and disease: Association or causation?' *Proceedings of the Royal Society of Medicine*, vol. 58, pp. 295–300.

Wark, M 2005, *Keeping a project journal*, Hunter–Central Rivers Catchment Management Authority, Tocal.

Young, B, Nichols, S and Norris, R 2006, 'Application of multiple lines and levels of evidence (MLLE) for addressing ecological questions of causality', Australian Society for Limnology 45th Annual Conference, 25–29 September, Albury, NSW.