

Chapter 18

Monitoring and evaluation

18.1 Introduction

Monitoring and evaluation (M&E) are important because they:

1. Make operational agencies more accountable to those they seek to help, as well as those who support them.
2. Demonstrate to donors, policymakers and practitioners that risk reduction works, thereby making a case for greater effort in this area.
3. Improve understanding of how DRR works in practice – including identifying problems and mistakes.

This chapter contains a short account of approaches to M&E of DRR projects and programmes, focusing on evaluation. Project monitoring in general is covered in standard manuals and should be part of all agencies' systems and training. Evaluation is one of the principal methods by which agencies seek to learn lessons and incorporate them into their work to improve future policy and programming. It also provides accountability to partners, beneficiaries and donors.

The range of M&E approaches and methods in development and relief has grown considerably over the years, as has the level of interest in the subject. This has partly been driven by criticism and donor pressure, but also by the desire to demonstrate success and improve performance. A growing body of work is providing agencies with better-informed guidance on M&E methods for development, DRR and emergencies. This is supported by initiatives such as the Active Learning Network on Accountability and Performance in Humanitarian Assistance (ALNAP) (www.alnap.org) and the electronic MandE information forum for development workers (www.mande.co.uk).

Assessment of a project or programme can focus on several different aspects:

- Inputs. These are the human, financial and technical resources deployed. Their effectiveness, cost-effectiveness and appropriateness can be assessed.
- Activities and processes. This covers the performance of tasks and factors affecting this.

- **Outputs.** These are the immediate results the project achieves (sometimes called ‘deliverables’).
- **Impact (or outcomes).** This is significant or lasting changes, brought about by a specific action or series of actions.¹

Similarly, the main distinctions between monitoring and evaluation can be identified:

- Monitoring usually addresses inputs, activities and outputs. Most monitoring systems are designed to meet the ongoing information needs of project managers and provide information for progress reports to donors. Evaluations focus on outputs and especially impact, and are intended for a wider audience within and outside the organisation.
- Monitoring is mainly descriptive. Evaluation is more analytical. Impact assessment is mainly analytical and concerned with longer-term outcomes.
- Monitoring should be regular and frequent, throughout the project. Evaluation is infrequent and can take place at any point in the project cycle (and after the project has ended).

Other terms used in this context are:

- **Review.** Reviews come somewhere between monitoring and evaluation. They supplement regular monitoring, taking place less frequently and providing an opportunity to identify key issues in programming. They usually form part of internal management systems, but reviews involving external stakeholders are not uncommon.
- **Audit.** Audits assess project and programme compliance with established regulations, procedures or mandates.

18.2 Planning and operation

M&E must be planned carefully, bearing in mind that no two projects are the same. Many agencies have developed generic evaluation criteria, which can be helpful. For example, the evaluation criteria set out by the Organisation for Economic Cooperation and Development’s Development Assistance Committee (OECD-DAC) are widely used in development projects, and often in humanitarian actions and DRR as well (see Box 18.1: OECD-DAC evaluation criteria). However, such frameworks and approaches should not be adopted thoughtlessly: they cannot be applied to every situation. They can be used to start discussions about what to evaluate and how to go about it, but the evaluation should be designed with the specific project in mind.

¹ C. Roche, *Impact Assessment for Development Agencies: Learning To Value Change* (Oxford: Oxfam/Novib, 1999).

Box 18.1 OECD-DAC evaluation criteria

Development

Relevance: The extent to which the aid activity is suited to the priorities and policies of the target group, recipient and donor.

- To what extent are the objectives of the programme still valid?
- Are the activities and outputs of the programme consistent with the overall goal and the attainment of its objectives?
- Are the activities and outputs of the programme consistent with the intended impacts and effects?

Effectiveness: A measure of the extent to which an aid activity attains its objectives.

- To what extent were the objectives achieved/are likely to be achieved?
- What were the major factors influencing the achievement or non-achievement of the objectives?

Efficiency: Measuring the outputs in relation to the inputs.

- Were activities cost-efficient?
- Were objectives achieved on time?
- Was the programme or project implemented in the most efficient way compared to alternatives?

Impact: The positive and negative changes produced by a development intervention, directly or indirectly, intended or unintended.

- What has happened as a result of the programme or project?
- What real difference has the activity made to the beneficiaries?
- How many people have been affected?

Sustainability: Measuring whether the benefits of an activity are likely to continue after donor funding has been withdrawn.

- To what extent did the benefits of a programme or project continue after donor funding ceased?
- What were the major factors which influenced the achievement or non-achievement of sustainability of the programme or project?

Box 18.1 (cont'd)

Humanitarian action

Relevance/appropriateness:

- Relevance: Assessing whether the project is in line with local needs and priorities (as well as donor policy).
- Appropriateness: Tailoring humanitarian activities to local needs, increasing ownership, accountability and cost-effectiveness accordingly.

Connectedness: The need to ensure that activities of a short-term emergency nature are carried out in a context that takes longer-term and interconnected problems into account.

Coherence: The need to assess security, developmental, trade and military policies as well as humanitarian policies, to ensure that there is consistency and, in particular, that all policies take into account humanitarian and human rights considerations.

Coverage: The need to reach major population groups facing life-threatening suffering wherever they are.

Efficiency: Measuring the outputs – qualitative and quantitative – achieved as a result of inputs. This generally requires comparing alternative approaches to achieving an output, to see whether the most efficient approach has been used.

Effectiveness: The extent to which an activity achieves its purpose, or whether this can be expected to happen on the basis of the outputs. Implicit within the criterion of effectiveness is timeliness.

Impact: the wider effects of the project (social, economic, technical, environmental) on individuals, gender and age groups, communities and institutions. Impacts can be intended and unintended, positive and negative, macro (sector) and micro (household).

OECD-DAC, *DAC Criteria for Evaluating Development Assistance* (Paris: OECD-DAC, 2000), <http://www.oecd.org/development/evaluation/49756382.pdf>; ALNAP, *Evaluating Humanitarian Action Using the OECD-DAC Criteria: An ALNAP Guide for Humanitarian Agencies* (London: ALNAP, 2006), <http://www.alnap.org/material/78.aspx>.

Evaluations can take many forms, including real-time evaluations, after-action reviews with communities, internal or self-evaluations by project staff and partners, and formal, externally-led evaluations. Evaluation guidelines sometimes divide them into three main kinds: summative (judging the merits and achievements of a project or programme), formative (to enhance project/programme learning, by understanding what worked and why), and developmental (to introduce new learning and ideas to organisations or the sector as a whole).² Mid-term evaluations are usually undertaken in longer projects. Impact evaluations long after the conclusion of the project are also valuable, but are rare.

The evaluation process should begin at the project design stage, when goals and objectives are set and logical or other results-based frameworks developed. Ideally, there should be a series of evaluations during and after the project, to permit longitudinal analysis, although this rarely happens. Evaluations should be scheduled at those points in the project where they can be most useful, principally at key moments for decision-making.

The purpose and methods of any monitoring exercise, review or evaluation should be clearly defined and agreed. Since it is almost never possible to assess everything, there must be some focus to the assessment, and its objectives must be realistic in relation to the resources that go into it. Thought should be given to such issues as:

- Indicators (see Section 18.7) – this is very important.
- Units of assessment. M&E can take place at individual, household, group, community, institutional, district and national levels. Even in a large project, it is important to get as close to the grassroots as possible: data can be collated subsequently.
- Sampling: sample size and sampling methods.
- Scope. This is conditioned by the project's coverage in terms of geographical area, hazards and risks addressed, and the number and types of vulnerable people assisted, as well as by factors such as remoteness, difficulty of access and security.
- Existing information sources. Most evaluations draw on external sources (e.g. government and other agencies' data sets and surveys) and internal sources (e.g. project documents) as well as field surveys. The quality and accessibility of external data is likely to vary. Agencies need to have knowledge management systems in place to identify and obtain relevant internal documents.
- Who should be involved in collecting, providing and discussing evidence (see also Section 18.4: Accountability and participation). The size, composition and skills of evaluation teams are important considerations.

2 M. Buchanan-Smith and J. Cosgrave, *Evaluation of Humanitarian Action: Pilot Guide* (London: ALNAP, 2013), pp. 33–34.

- **Scheduling.** Reviews and evaluations should be scheduled at an appropriate point in the project's lifetime, or after it has ended, and at suitable times of year (i.e. not at periods when communities are very busy, such as harvest time, or when weather conditions are difficult). When in the field, researchers must find appropriate times of the week or day for talking to the different beneficiaries.
- **Tools and methods.** In the field, these may comprise formal surveys, structured or semi-structured individual and group interviews, group discussions such as focus groups and workshops, direct observation, community mapping, seasonal calendars, timelines, problem/solution trees, other participatory learning and action methods and case studies. Each method brings its own advantages and drawbacks. Project evaluations generally use several methods. The methods adopted must be appropriate to what is being assessed and the resources available to carry out the assessment.
- **Matching inputs and outputs.** The evaluators must have enough time and resources to carry out the proposed activities and achieve the outputs required. Effective M&E also requires organisational capacity to support it: systems, financial resources and specialists in data management, analysis, reporting and M&E training.
- **How the findings will be reported back to all the stakeholders concerned, and how they will be acted upon.** This is often neglected.

Clear terms of reference are vital. Many problems with evaluations stem from a failure to achieve this clarity and reach agreement on it. Sufficient time should be set aside for this. Even the best plans can break down when confronted with reality in the field, so flexibility is essential. Good planning should allow for this.

Often, evaluation teams are not given the time or resources to do their work thoroughly. Overcrowded schedules are common. This limits time for preparation and in the field, forcing evaluators to place too much reliance on what may be very selective field evidence, on agency documents that may be incomplete or unavailable and on interviews in head offices. In consequence, many evaluations are little better than snapshots of an initiative, coloured by chance encounters and personal views. Experienced evaluators can compensate for this to some extent by drawing on their skills in identifying and gathering key data and their knowledge of similar initiatives elsewhere, but if they rely too much on their general knowledge they may miss features that are distinctive to the programme or project in question. However, snapshots can be useful. Small-scale or rapid assessments do provide valuable insights in some cases, especially when focused on a distinct aspect of risk reduction (as in Case Study 18.1: Reporting on disaster response).

Case Study 18.1 Reporting on disaster response

On 11–12 November 2002 a cyclone warning was issued along the coast of the Bay of Bengal. A relatively weak cyclone struck, with high winds and heavy rain in several places. The Orissa State Branch of the Indian Red Cross used the event to assess the effectiveness of its disaster preparedness work. The initial assessment was based on telephone calls from local voluntary coordinators and emergency team members in eight locations. These conversations focused on the following:

- When the cyclone warning was received, and from which source(s).
- Actions taken by local disaster preparedness teams.
- Actions taken by villagers.
- Details of the event (wind speed, condition of the sea, rainfall) and its impact.

The phone calls provided plenty of local detail. Using this, it was possible to build up a picture of the situation on the ground and actions taken almost as they happened, the effectiveness of warning and response mechanisms and factors affecting them, and variations between the locations. The phone call method was not seen as a substitute for field surveys, but it would not have been possible to carry out such surveys immediately after the event.

Orissa State Branch, Indian Red Cross Society, 'Actions by 8 Red Cross Cyclone Shelter Communities in Orissa during Cyclone Warning (Nov. 11 to 12, 2002)', mimeo, 2002.

18.3 Sustainability and theory of change

It can be difficult to judge a project's long-term sustainability and replication, but this can often be inferred from other evidence. For example, DRR initiatives are more likely to be sustainable where extensive time and effort have gone into preparatory work with communities, partners and other local and national actors. Another indicator is the level of stakeholder contributions of financial and other material and human resources to the project (on the assumption that sustainability is linked to the degree of local ownership).

In community-based projects, the strength of community organisation is central to sustainability. Evaluations often place great emphasis on the creation or revival of local groups such as disaster management committees. The mere existence of such groups is a

weak indicator of their capacity, whilst attitudinal analysis may only demonstrate short-term enthusiasm. Evidence of group activity should be collected (e.g. risk assessments, preparation of emergency plans, building of mitigation structures). The frequency, nature and quality of such activities and the degree of community involvement can be monitored and evaluated. Evaluators should also consider external factors that may affect sustainability, such as changes in official policy, staff turnover and economic changes.

Whatever their focus, evaluations – and projects – should be based on a ‘theory of change’. Theory of change is understood and approached in different ways, but the key idea is that individual initiatives should be underpinned by broader thinking about how change happens generally, as well as the overall, longer-term changes the initiative itself seeks to achieve (or to contribute to, where the desired changes are very long-term). These ideas need to be acknowledged, debated and made explicit by project participants, and projects should be clear about how they contribute to change. This generates a shared vision or rationale at the start, provides clarity about roles and ways of working during the project and makes it easier to assess and communicate findings at the end.³

Formal planning, monitoring and evaluation tools, such as logical and results-based frameworks (see Section 18.9: Identifying cause–effect links) can capture some of this, but they tend to be linear and compartmentalised, and may be less effective in complex programmes or in explaining how individual projects relate to the larger, more complex socio-ecological and socio-political systems in which they are located. Uncertainty, which is an important element in risk assessment and long-range planning for sustainable development, DRR and climate change adaptation, is another issue that conventional M&E approaches may not be able to capture adequately.

18.4 Accountability and participation

It is best to approach M&E as a mutual learning process for all involved, not merely as an information-gathering exercise. This encourages flexibility, openness and debate. The principles of accountability to vulnerable people outlined in Chapter 11 are very important here. Communities’ views should be central to evaluation, and communities (or beneficiaries) should be able to take an active part in the evaluation process. Participatory evaluation enables the voices of project stakeholders, particularly beneficiary communities and vulnerable groups, to be heard, draws on their local knowledge, stimulates dialogue and

³ C. James, *Theory of Change Review: A Report Commissioned by Comic Relief* (London: Comic Relief, 2011), http://www.theoryofchange.org/pdf/James_ToC.pdf; I. Vogel, *Review of the Use of ‘Theory of Change’ in International Development* (London: DFID, 2012), <https://www.gov.uk/government/news/dfid-research-review-of-the-use-of-theory-of-change-in-international-development>.

Box 18.2 Differences between theory of change and logical frameworks

Theory of change:

- Captures the complexity of change much better – it is less linear and can take into account relationships and the interdependence between different programme elements.
- Acknowledges an initiative's starting point (captures existing change processes).
- Is broader than just a project: it takes into account programme and non-programme factors affecting change.
- Focuses on change and how to get there, rather than programme components.
- Is simpler and more flexible – more 'organic' and less 'mechanistic'.
- Is not rigid in its terminology and so is more free-flowing.
- Is an exercise or process, rather than a tool.
- Enables organisations to think about their work and their organisation more deeply.
- Is more interactive and fun.

C. James, *Theory of Change Review: A Report Commissioned by Comic Relief* (London: Comic Relief, 2011), http://www.theoryofchange.org/pdf/James_ToC.pdf.

mutual learning and creates wider 'ownership' of the evaluation's findings. However, many M&E systems are still top-down, designed to extract information from the field to give to headquarters staff and donors. Collecting data solely for external use can undermine the participatory process.

Beneficiary participation in M&E can take various forms. In some projects, it may be no more than providing information to review or evaluation teams, but this is too limiting. Beneficiaries should be involved in planning the assessment (including selecting indicators), providing information on what was and was not achieved, analysing and verifying the results and making decisions about future activities. Findings should always be fed back to communities. The needs of communities in this regard may differ from those of outside agencies, and the targets, indicators and priorities developed by communities may differ

considerably from those of agency staff. Adopting participatory approaches does not prevent the use of more formal data collection methods: these can complement or validate information gathered in a participatory way. Methods should be selected according to their usefulness in helping to understand impact.

Participatory methods such as those described in Chapter 6 are valuable in allowing beneficiaries to express their views. Standard participatory learning and action exercises can yield valuable information. Since it is never possible to involve everyone, careful thought must be given to ensuring that those who are consulted are representative of the range of groups concerned, paying particular attention to the most marginalised as well as people who may have dropped out of the project. Some evaluations pick up the views of similar people who were not involved in the project as a kind of ‘control group’.

Participatory impact assessment focuses on a project’s impact on beneficiaries’ lives rather than measuring project performance. The importance of identifying who benefits from a DRR initiative, and who does not, cannot be overemphasised. Evaluators should never assume that benefits are spread evenly across a community. They should assess beneficiary communities’ socio-economic characteristics carefully, considering gender issues and people who are vulnerable due to other factors, such as poverty, ethnicity, age and disability, as well as the influence of local power relationships.

Beneficiaries are one group of stakeholders. Project staff are another. NGOs and other local institutions, local and national government officials, and, where appropriate, international donor agencies and other kinds of organisation (e.g. the private sector) should be consulted if they have been involved in the project, are affected by it or have some influence on its outcome. It can be difficult to reconcile the views of such diverse groups. This makes it all the more important to be clear from the start about what M&E is designed to look at. Meetings should be held to discuss and explain this. Where stakeholders have different priorities and perspectives, this should be made explicit at the start to avoid misunderstandings later.

Evaluations are often funded by donors or in some way linked to ensuring continued donor support. In such circumstances the ideal of M&E as mutual learning may be hard to sustain. Many of those involved will be tempted to overstate the positive features of their project and downplay the negative ones. They may be defensive about their work, fearing that evaluation teams are searching for faults and problems. Community members may only tell evaluators what they think they want to hear.

18.5 Assessment teams

Participation and accountability are significant factors to bear in mind when forming assessment teams. The balance between internal and external assessors is an important consideration. Evaluations may be carried out by external specialists, local staff or local people, working separately or in mixed teams. Both internal and external evaluations aim to learn lessons, but external evaluations, which provide a more detached, objective perspective, also make an important contribution to accountability. There are no fixed rules: the appropriate size and mix of evaluation team depends on the specific project. However, all teams should have the appropriate technical skills, gender balance and local participation.

Involvement of a range of people makes it more likely that an evaluation's lessons will be shared and its findings acted upon. Unfortunately, external specialists – mostly men – often dominate teams evaluating DRR and humanitarian aid initiatives, and it is still common to have projects evaluated by a single external consultant. Whilst it is useful to have the added objectivity of an outsider's view and the experience of a well-travelled evaluator, there is a danger that somebody new to the project will not understand all its complexities. This danger is accentuated by the limited time usually allocated to evaluators.

The purpose of the evaluation offers some guidance on the balance of the evaluation team. If the main purpose is lesson learning, it makes sense to involve more internal staff; if it is accountability, the independence of external evaluators becomes more important. In practice, however, most evaluations aim at lesson learning and accountability. There is a lot of discussion in the literature about the appropriate skills mix in evaluation teams. Again, there are no fixed rules about this. Some people feel that a wide range of technical skills is essential; others maintain that experience in evaluation methods is more important. In some kinds of DRR project technical expertise may be valuable, be it in science, engineering, architecture, nutrition, economics or the social sciences. Evaluators need to be able to use relevant data types (e.g. quantitative or qualitative) and collection methods. Knowledge of local geography, society, cultures and institutions is also important.

18.6 Baseline data

A baseline is the measurement of conditions at the start of a project, against which subsequent progress can be assessed. Ideally all projects should have baseline studies to help them set their objectives and indicators of achievement. In practice, this happens far less than it should, leaving many evaluators struggling to find adequate measures of success.

In DRR programmes, a risk or vulnerability/capacity analysis (VCA) should provide good baseline data to guide planning and interventions. However, considerable resources are required for a comprehensive analysis, and time, resource and capacity constraints in

Case Study 18.2 **Establishing baselines**

Project assessments of its community-based DRR programmes by the Myanmar Red Cross include initial baseline surveys intended to provide community profiles and an understanding of local perceptions of risk and disasters, and to identify the coping and preparedness strategies used by households and communities. Baseline surveys are carried out by community M&E teams comprising between 12 and 15 local Red Cross volunteers. Team members are given a two-day basic training course in M&E concepts and survey methods, and are supported in the field by a specialist M&E officer. A sample of 10–20% of households is interviewed using a semi-structured questionnaire, and the data collected is entered onto a computer for analysis using specialist software: all of this work is done by the volunteers. The M&E officer then writes a report based on the data and analysis. End-of-project surveys are carried out by the same teams, using the same sets of questions as the baselines, in order to compare and measure change. Key respondents are also interviewed to add a further perspective on the project's impact.

Baseline and End-line Studies as Means To Analyze the Effectiveness of Community-based Disaster Risk Reduction Activities (Nay Pyi Taw: Myanmar Red Cross Society, undated).

evaluations make it difficult to collect and analyse the range of data required. VCA is also a relatively new approach for many project staff. Until they acquire greater confidence in the techniques required, they may be reluctant to use it in evaluations, especially where the findings will go before senior staff or funding agencies.

18.7 Indicators

Evaluators normally look for a range of indicators that will give a balanced view of a project's achievements and contribution towards its objectives: these should be easy to understand, by communities as well as implementing organisations. Indicators can be qualitative, quantitative or a mixture of the two, but in general they should try to be both SMART (specific, measurable, attainable, relevant and time bound) and SPICED (subjective, participatory, interpreted, cross-checked, empowering and diverse): see Table 18.1. Remember that the indicators that are easiest to measure are not necessarily the most useful for analysis.

This sounds simple on paper, but in practice it is more complicated. Questions to be asked regarding the practicality of indicators include:

Table 18.1 SMART and SPICED indicators

Smart	Spiced
<p>Specific: Indicators should reflect those things the project intends to change, avoiding measures that are largely subject to external influences.</p> <p>Measurable: Indicators must be defined precisely so that their measurement and interpretation are unambiguous. They should give objective data, independent of who is collecting the data. They should be comparable across groups and projects, allowing change to be compared and aggregated.</p> <p>Attainable: Indicators should be achievable by the project and therefore sensitive to the changes the project wishes to make.</p> <p>Relevant: It must be feasible to collect data on the chosen indicators within a reasonable time and at a reasonable cost. Indicators should be relevant to the project in question.</p> <p>Time-bound: Indicators should describe by when a certain change is expected.</p>	<p>Subjective: Informants have a special position or experience that gives them unique insights which may yield a very high return on the investigators' time. In this sense, what may be seen by others as anecdotal becomes critical data because of the source's value.</p> <p>Participatory: Indicators should be developed together with those best placed to assess them. This means involving a project's ultimate beneficiaries, but it can also mean involving local staff and other stakeholders.</p> <p>Interpreted and communicable: Locally defined indicators may not mean much to other stakeholders, so they often need to be explained.</p> <p>Cross-checked and compared: The validity of assessment needs to be cross-checked, by comparing different indicators and progress, and by using different informants, methods and researchers.</p> <p>Empowering: The process of setting and assessing indicators should be empowering in itself and allow groups and individuals to reflect critically on their changing situation.</p> <p>Diverse and aggregated: There should be a deliberate effort to seek out different indicators from a range of groups, especially men and women. This information needs to be recorded in such a way that these differences can be assessed over time.</p>

C.Roche, *Impact Assessment for Development Agencies: Learning to Value Change* (Oxford: Oxfam/Novib, 1999), pp. 48–49.

- Measurability. Is the indicator measurable? Is it sufficiently sensitive to an improvement or deterioration in conditions?
- Ease and cost of collection. How easy is it to obtain the information required? How costly will this be? Can the community participate? Are relevant data already collected?
- Credibility and validity. Are the indicators easy to understand, or will people argue over what they mean? Do they measure something that is important to communities as well as implementing organisations?
- Balance. Do the selected indicators provide a comprehensive view of the key issues?
- Potential for influencing change. Will the evidence collected be useful for communities, implementers and decision-makers?⁴

Even with this guidance in mind, it is very rare to find all the evidence one wants. Indicators are *indicators*: they are not necessarily final proof. In some cases it will not be possible to measure change directly, with clear and unambiguous indicators. It is often necessary to identify relative or approximate changes instead, using indirect or ‘proxy’ indicators.

Part of the process of collecting baseline information should be to identify those indicators that will be most valid for M&E. However, experience as the work progresses may highlight other issues and require changes to the project. Some indicators may have to be modified or new ones will emerge, which makes it important to be flexible. Monitoring methods should be designed to pick up these issues so that decisions can be made. Where baseline data are lacking (which is often the case), or previously identified indicators are difficult to assess or simply irrelevant, the baselines may have to be reconstructed (e.g. from project documents, interviews with key informants and data from other organisations) or new indicators must be developed. In practice this happens quite often, but the process must be managed carefully to avoid confusing or misleading stakeholders; an open, participatory approach is needed, and the aim should be to achieve the highest possible level of consensus.

Evaluations usually combine qualitative and quantitative data. Both types are valuable, in different ways. Quantitative indicators are often used to assess progress towards stated targets (e.g. the number of hazard-resistant structures built or community disaster preparedness committees established). Numbers alone cannot measure quality or effectiveness, although they can be proxy indicators for this. Qualitative data are often used in DRR evaluations. Typically they are collected from stakeholders through workshops, focus

⁴ L. Nosen, ‘Hazard Mapping and Risk Assessment’, in ADPC (ed.), *Proceedings, Regional Workshop on Best Practices in Disaster Mitigation, 24–26 September 2002, Bali, Indonesia* (Bangkok: Asian Disaster Preparedness Center, 2002), pp. 83–84 <http://www.adpc.net/audmp/rllw/default.html>.

Table 18.2 Examples of data collection methods and their application

Method	Examples of application
Formal surveys of beneficiaries and other stakeholders (can also be generated by interviews and group discussions)	<ul style="list-style-type: none"> • Survey of builders and occupants of hazard-resistant housing to ascertain application of skills and increased security • Household survey on food production, availability, consumption and marketing to identify patterns and shifts in vulnerability
Structured and semi-structured interviews with staff, partners, beneficiaries and others	<ul style="list-style-type: none"> • Individual interviews building up a picture of the level of understanding of the project, agency–community working relationships, effectiveness of coordination mechanisms and outcomes of DRR interventions
Group discussions, especially with beneficiary communities (e.g. participatory workshops, focus groups)	<ul style="list-style-type: none"> • Beneficiary workshop to identify and assess benefits of particular DRR interventions and unforeseen impacts • Expert workshop to assess potential effectiveness of new DRR methods or approaches • Feedback workshop with beneficiaries and other stakeholders to test/confirm evaluation findings
Rapid assessments	<ul style="list-style-type: none"> • Post-disaster telephone or field survey to indicate effectiveness of warning and response mechanisms and factors affecting them
Direct observation and visual surveys	<ul style="list-style-type: none"> • Visual surveying of structural mitigation measures to determine quality of design and workmanship, take-up of technologies or techniques (disaster resilience inferred from this or assessed through post-disaster surveys) • Observation of coping strategies and other risk-reducing behaviour – before, during and after disasters

Table 18.2 (cont'd)

Method	Examples of application
Case studies	<ul style="list-style-type: none">• Personal or group accounts of use of skills, materials and organisational capacity acquired from disaster management training courses, during subsequent events
Simulations	<ul style="list-style-type: none">• Group simulation or exercises (table-top or field) of disaster management activities or responses to disaster events, to test plans, skills, equipment, etc.
Documentary evidence	<ul style="list-style-type: none">• Content analysis of educational material on risk reduction and management produced by project• Quantitative and qualitative data about project delivery, effectiveness, impact and costs, from project documentation• Secondary data collection to complement or validate information collected by the evaluators in the field

C. Benson and J. Twigg with T. Rossetto, *Tools for Mainstreaming Disaster Risk Reduction: Guidance Notes for Development Organisations* (Geneva: ProVention Consortium, 2007), http://www.preventionweb.net/files/1066_toolsformainstreamingDRR.pdf.

groups or semi-structured interviews. They can provide good measures of achievement and impact, and reveal insights into processes and attitudes. Participatory approaches tend to produce a good deal of qualitative information. Some examples of data collection methods and their application are shown in Table 18.2.

18.8 Outputs or impact?

M&E manuals often speak of ‘impact’ and ‘process’ indicators. Impact indicators, which can be both quantitative and qualitative, measure changes that occur as the result of project activities. Conventional M&E methods usually focus on positive impacts. Few initiatives are without some negative impacts, although in most projects there is a reluctance to review these. All partners in a project should be open about the importance of identifying negative impacts and groups

that have been overlooked or excluded. This requires a high degree of trust between those involved in the project, which may be difficult to achieve owing to the unequal relationship between poor communities and external organisations bringing in resources.

Process indicators measure the implementation of project activities, and are usually quantitative. They often act as proxy indicators of impact for DRR interventions, especially where the hazards concerned are infrequent. Actions during a project can be used as indicators of potential effectiveness. In a community disaster preparedness project, for example, process indicators might include recruiting, training and establishing a community disaster management team, organising public meetings to identify threats and the most vulnerable households, building relevant structures and holding regular evacuation drills.

In practice agencies are more comfortable with indicators of output rather than impact (especially quantitative indicators), and it is common for evaluations to come up with the kinds of output indicators that merely quantify the measures taken by a project (e.g. the number of volunteers trained or public education leaflets distributed). Evaluations tend to be short-term studies, usually carried out at the end of a project, when it is too soon to assess its longer-term consequences. Post-project impact assessments are rarer and there is a shortage of genuinely long-term studies. Published case studies of well-regarded initiatives usually appear at a relatively early stage in the project or are based on short-term evidence. The exceptions tend to be drought/food security initiatives: these demonstrate that projects' impact can be judged only over a period of several years; they also reveal the extent of rethinking and modification that takes place even in successful projects.

DRR sometimes presents problems of evaluation because of what can be called its 'reverse logic': i.e. the success of an initiative is that something – the disaster or loss associated with it – does *not* happen. Nevertheless, evidence from subsequent hazard events and the response to them is a good indicator of the impact of some types of DRR intervention, such as the effectiveness of early warning and response systems, and the resilience of houses and infrastructure.

Structural/physical mitigation measures are relatively easy to assess. The quantity and quality of, for example, embankments, flood shelters, earthquake-resistant houses and soil and water conservation structures can be assessed visually, as can the extent to which alternative technologies or techniques are adopted. Judgement about the quality of such technical innovations serves as a proxy indicator for their impact – i.e. their resilience to actual hazard events. Non-structural measures involving changes in attitudes, skills, organisation or awareness are much more challenging. Proxy indicators of impact can be picked out, but they are less certain than those for physical change. For example, interviews or discussion groups can reveal how interventions have changed a community's attitudes towards risk, but only allow us to estimate how that community will actually behave when confronted with a disaster.

Case Study 18.3 Evaluating the impact of rainwater harvesting

In 1997 the NGO Intermediate Technology (now Practical Action) commissioned an independent evaluation of a rainwater harvesting initiative in Kenya that had begun more than ten years before. The evaluation was based on project documentation (including local partners' monitoring records), interviews with project and partner staff, five group discussions with beneficiaries (104 people in total), individual interviews and field observation. The discussion groups and individual interviews were based on PRA techniques. The evaluation covered a range of issues, including impacts on sorghum production, diets and household wealth, gender, land tenure and the environment.

Much of the evidence was qualitative. To obtain relative data on sorghum yields and constraints on sorghum production, the evaluators used ranking and proportional piling, in which individuals were asked to place stones in separate piles to indicate amounts. Data on crop yields was gathered from various sources, including project records, discussions with project staff and the assessments of interviewees. This was compared with data from previous project reviews and workshops.

C. Watson and B. Ndung'u, 'Rainwater Harvesting in Turkana: An Evaluation of Impact and Sustainability', mimeo (Nairobi: ITDG, 1997).

Given these challenges, the need for triangulation and cross-checking of different types of evidence is clear. This is particularly important for qualitative data, where evidence may be more subjective. Triangulation of interview or focus group data can also identify differences in partners' aims and expectations. Good impact evaluations should be wide-ranging in their search for relevant signs of increased resilience to risk, as well as objective about the quality of the evidence collected. Case Study 18.3 (Evaluating the impact of rainwater harvesting) is an example of this. In the field, direct observation is a useful way of identifying discrepancies between what people say and what they do, although evaluators do not always have enough time to do this.

18.9 Identifying cause–effect links

Analysis of the relationship between process (activity and output) indicators and outcome or impact indicators helps to understand cause–effect links, often referred to as ‘attribution’ in M&E guidance.

Many factors combine to make people vulnerable and create situations of risk. No project intervention can address all of these factors, and all projects will be influenced by them. This influence must be understood in order to assess a project’s achievements. To what extent are particular changes due to the project itself, or to local actors, external agencies and other factors? It can be difficult to make a judgement here, particularly when evaluating long-term impact.

Good risk reduction work should comprise a range of activities: organisational, educational, structural and socio-economic. Activities are meant to be mutually reinforcing. For example, training in safe building techniques should be complemented by regulation of land use and the setting and enforcement of building standards, as well as by measures to address the economic and social pressures that force poor people to live in flimsy housing in hazardous locations. Where risk reduction adopts such a broad approach, with numerous interlocking elements, how can one assess the results arising from one particular type of intervention against another? It may be impossible to identify specific links between cause and effect. Consequently, how can one set priorities for intervention?

Some project evaluations or assessments have used control groups for comparative purposes, although DRR and particularly humanitarian response agencies are sometimes uneasy about studying at-risk groups that the organisation is not attempting to protect. There are also methodological challenges with this approach: no two communities are exactly the same, which makes comparison difficult. The method is better at demonstrating the basic point that DRR interventions can bring benefits (by comparing communities that have been assisted against those that have not) than at assessing the most effective types of intervention to adopt. However, it can be useful. Some evaluations seek the views of community members not involved in projects, usually to identify reasons for non-participation. Talking to groups that have dropped out of a project can also provide valuable information about the way the project was implemented.

Some agencies specifically investigate external influences when assessing projects: this at least puts evaluation findings into context, even if it often cannot demonstrate specific cause-and-effect links. Triangulation of different data sets and sources is also helpful in isolating particular factors affecting success or failure. In most cases the sources and types of information will vary. In particular, there will be a mixture of quantitative and qualitative information. Using different stakeholders or assessors to review the same issue can reveal similarities and differences; here it is very important to consider the views of differently vulnerable groups.

The problem is reduced wherever evaluators can focus on specifics. Assessment of disaster preparedness and response measures tends to be simpler: for example, warning and evacuation procedures can be tested through practice drills as well as by events (there are examples of evaluation teams observing such drills). It is also relatively easy to isolate for analysis different elements in the preparedness-response system. Responses to early warnings have been studied on many occasions, throwing light on community attitudes and the effectiveness of early warning systems. Such knowledge has supported the development of sophisticated methods for evaluating the condition of early warning systems.

Projects that have clear objectives and targets can develop a hierarchy of indicators that link process to impact and thereby make M&E more coherent. Results-based frameworks, such as logical frameworks, which are used in project design, should already provide a hierarchy, helping evaluators to form judgements at all levels (activity, output, outcome, impact). However, M&E systems also need to be sensitive to changes and impacts that are due to a project, directly or indirectly, but which are unexpected and unplanned for. This means looking beyond formal, linear planning frameworks.

The Outcome Mapping and Most Significant Change methods move away from a focus on project results to explore how interventions contribute to change in wider, more complex and uncertain contexts. Outcome Mapping looks at changes in the behaviour, relationships and actions of the groups and individuals that a project works with (which may or may not be direct consequences of the project) and considers how the project and other factors contribute to that change process. Most Significant Change is a form of participatory M&E that works without predefined indicators, in which community members and field staff collect, discuss and analyse changes.⁵ These methods are good at capturing unforeseen changes and building up a more complete picture of change overall.

18.10 Cost–benefit analysis

Disasters cause huge economic losses through the damage and destruction of infrastructure, housing, crops, natural resources and other livelihood assets, and through lost production and employment. Relief and reconstruction are also costly. The case for DRR is often made in economic terms, by using cost–benefit analysis (CBA) to justify investment in protection generally, or in particular types of risk-reducing interventions. In addition, agencies and donors want evidence that their projects and programmes are delivering good value for

⁵ S. Earl, F. Carden and T. Smutylo, *Outcome Mapping: Building Learning and Reflection into Development Projects* (Ottawa: International Development Research Centre, 2001), <http://www.outcomemapping.ca/resource/om-manual>; Outcome Mapping website: <http://www.outcomemapping.ca/resource>; R. Davies and J. Dart, *The 'Most Significant Change' (MSC) Technique: A Guide To Its Use*, 2005, <http://www.mande.co.uk/docs/MSCGuide.pdf>; Most Significant Change web page: <http://mande.co.uk/special-issues/most-significant-change-msc>.

money. Evidence produced for the costs and benefits of DRR frequently shows that investment brings greater benefits than costs. However, every case is different, specific to location, hazard and type of intervention, and hence cost–benefit ratios may vary considerably (and may even be unfavourable).⁶

CBA has long been used as an economic assessment tool in large-scale mitigation projects, especially involving the protection of infrastructure and other physical structures. It is now being used increasingly in planning and, in particular, evaluating local-level DRR initiatives.⁷ As yet there is no common or standard methodology for this, and a variety of approaches have been used.

At the start of a project, CBA can be used alongside VCA to identify which interventions will obtain the best benefits with the resources available. However, it is not a routine part of project appraisal or evaluation in small-scale projects. This is partly because the principal

Case Study 18.4 Cost–benefit assessment of a DRR project

The NGO Tearfund commissioned a cost–benefit analysis of a DRR and food security programme in Malawi, which had been running for four years and covered 53 villages. The programme’s key components were crop diversification, soil and water conservation and provision of livestock (goats) that could withstand drought. The analysis was based on quantitative estimates of the programme’s benefits, in terms of increased crop and livestock production, continued education (by preventing drop-out due to hunger and inability to pay school fees) and a reduction in hunger-related malnutrition and mortality. It was calculated that the programme had delivered \$24-worth of benefits for every \$1 invested.

C. Cabot Venton and J. Siedenburg, *Investing in Communities: The Benefits and Costs of Building Resilience for Food Security in Malawi* (Teddington: Tearfund, 2010), <http://tilz.tearfund.org/~media/files/tilz/research/investing%20in%20communities%20web.pdf>.

⁶ C. Shreve and I. Kelman, ‘Does Mitigation Save? Reviewing Cost–Benefit Analyses of Disaster Risk Reduction’, *International Journal of Disaster Risk Reduction*, vol. 10, part A, 2014, <http://www.sciencedirect.com/science/article/pii/S2212420914000661>.

⁷ For examples, see C. Cabot Venton and P. Venton, *Disaster Preparedness Programmes in India: A Cost–Benefit Analysis*, Network Paper 49 (London: ODI, 2004), <http://www.odihpn.org/hpn-resources/network-papers/disaster-preparedness-programmes-in-india-a-cost-benefit-analysis>; T. H. Tuan and B. D. Tinh, *Cost–Benefit Analysis of Mangrove Restoration in Thi Nai Lagoon, Quy Nhon City, Vietnam* (London: International Institute for Environment and Development, 2013), <http://pubs.iied.org/10644IIED.html>; D. Willenbockel, *A Cost–Benefit Analysis of Practical Action’s Livelihood-Centred Disaster Risk Reduction Project in Nepal* (Brighton: Institute of Development Studies, 2011), <http://www.ids.ac.uk/idsproject/cost-benefit-analysis-livelihoods-centred-disaster-risk-reduction>.

local actors, NGOs and CBOs, are unwilling to give too much weight to purely quantitative features of complex socio-economic processes, partly because of their lack of familiarity with the methods, and partly because of the perceived difficulty of carrying out this kind of analysis. CBAs are generally quantitative, using data from primary and secondary sources, but they can also incorporate qualitative aspects, especially when carried out at community level as part of a participatory process, or to explore quantitative findings more extensively. For instance, ISET-Nepal has facilitated ‘shared learning dialogues’ with local stakeholders to identify and discuss their perceptions of the costs and benefits of different flood mitigation interventions and develop appropriate risk reduction strategies.⁸ Data gathering does not necessarily require considerable extra resources or technical capacity, depending on the data available or the level of analysis. Where data are limited quantitative assessment may not be possible, or would provide misleading results.

There are several challenges and issues regarding the use of CBA in risk reduction. It is difficult to assess the human and economic impact or cost of disasters. Data and methods have improved over the years but remain unreliable, especially in low-income countries. Estimates of economic impact generally focus on direct costs, and it is more difficult to assess indirect and secondary costs (see Box 18.3: Economic costs of disasters). Assessment

Box 18.3 Economic costs of disasters

The economic costs of disasters are usually divided into three kinds: direct, indirect and secondary.

- Direct costs relate to the capital cost of assets (e.g. buildings and other physical infrastructure, raw materials, crops) destroyed or damaged.
- Indirect costs are the damage to the flow of goods and services (e.g. lower output from factories destroyed or damaged, loss of sales income due to damaged infrastructure, costs of having to buy materials or services from elsewhere, medical expenses, lost productivity).
- Secondary effects are the short- and long-term impacts on overall economic performance (e.g. deterioration in external trade and government budget balances, reallocation of planned government spending, increased indebtedness, changes in income distribution patterns, changes in the scale and incidence of poverty).

⁸ A. Dixit, A. Pokhrel and M. Moench, *Costs and Benefits of Flood Mitigation in the Lower Bagmati Basin: Case of Nepal Terai and North Bihar* (Kathmandu: ISET-Nepal and ProVention Consortium, 2008), http://www.preventionweb.net/files/81116_WP206highres1.pdf.

of direct losses is harder where a large proportion of losses are uninsured, and it is difficult to calculate less direct costs such as loss of income when a significant proportion of economic activity takes place in the informal sector.

A focus on economic costs and benefits addresses only one aspect of people's vulnerability to disasters. One of the main criticisms of CBA in DRR is that it values costs and benefits in purely monetary terms. In the case of physical structures (e.g. homes, infrastructure, public buildings) and economic aspects (e.g. employment, crops and livestock, savings) these calculations are relatively straightforward. It is much more difficult to quantify less tangible aspects (e.g. the natural environment, social and psychological issues) and many CBAs do not pay enough attention to them. Projects with clear monetary benefits may be selected over those which may be equally beneficial, but whose results are not so easily quantified: this is problematic for community DRR, which typically includes a mixture of 'hard' and 'soft' measures. Other challenges include incorporating uncertainty and trends into assessments. Calculating the probability and extent of a hazard's occurrence and impacts can be difficult, especially at local level and where there are data gaps and deficiencies. CBA is better at assessing shorter-term outcomes than longer-term trends, where there is a much higher level of uncertainty. Climate change adds another level of complexity.

There are also ethical concerns, the main one being that many people object in principle to assigning a monetary value to human life. Another is that conventional CBA does not consider the distribution of costs and benefits within communities (in other words who gains from DRR measures, and who loses out). Additional qualitative assessment may be needed to identify the impacts on different households, social groups, businesses and institutions.

18.11 Using M&E findings

M&E is of little value if it does not lead to improvements in agencies' work to reduce risk. M&E reports are potentially very useful documents. They enable practical lessons to be learned and applied within and across programmes and regions. They feed into strategic planning by providing a basis for discussion about better practice and policy change. They also contribute to institutional memory, which is important in organisations that suffer from rapid staff turnover. Good-quality presentation is essential here: no matter how good the evidence and analysis they contain, reports will not inform and influence if they are not well written and presented.

Evaluation should be embedded within an organisation's systems and regular practice to ensure that learning takes place. In reality, many agencies are poor at absorbing the lessons from evaluations, with the result that the same problems recur. Too often, the review or evaluation report is filed away to be acted upon later, but then forgotten amidst the pressure of work. Many organisations have poor information storage and retrieval systems, making

it very difficult to find documents, and feedback mechanisms are weak. Few staff have sufficient time to reflect on the lessons from individual projects, and fewer still are able to consider what can be learnt from several projects and countries. Overwork and pressures of work, which are common among staff in DRR agencies, prevent clear thinking and innovation. Knowledge management and learning systems need to be given higher priority and more resources in most organisations. Plans for sharing and using results and findings, in the field and across the organisation, should be built into the evaluation process from the start. These should be based on consultations with potential users of the evaluations.

Transparency in M&E is a key element in making operational agencies more accountable. Evaluation processes should be as open as possible, and their results should be made widely available, particularly to project stakeholders (who should also be consulted before reports are submitted, for clarification and confirmation). However, there is still much to be done here. The widespread failure to share and publish DRR evaluations means that practitioners are unable to learn lessons from each other and so are frequently reinventing the wheel. It also runs counter to the principle of accountability that agencies claim to follow. There is a particular reluctance to document mistakes and share their lessons. In some cases, joint reviews by agencies could be carried out to encourage mutual learning, knowledge sharing and transparency. Participatory M&E creates a sense of 'ownership' of the final product among stakeholders, which greatly increases the likelihood that lessons will be noted and acted upon.