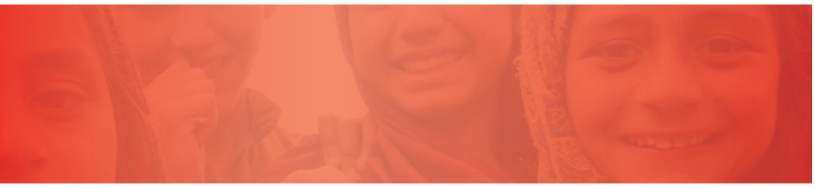
Save the Children

The Open University

Monitoring, Evaluation, Accountability and Learning (MEAL)

5 Baseline and evaluation design and management

Keywords: Evaluation, baseline, evaluation approach, evaluation design, evaluation management, systematic assessment, attribution, contribution, gross effect, net effect, recommendations, use of evaluation findings and learning



Introduction

How do you know if a given intervention is working? How can you measure and demonstrate if it is producing changes in a systematic manner? Are there different ways of doing so? This session will answer this and other similar questions and will provide you with useful skills to manage or actively participate in the process of conducting a baseline study, needs assessments, project evaluations and real-time reviews.

Indeed, these questions and processes are key to improving the effectiveness and quality of our programmes. In our dual mandate Quality Framework explicit components related to evaluation and learning are stated, with the aim of improving programme quality and measuring and demonstrating the impact for children.

Learning Outcomes for this session

Knowledge and Understanding

When you have studied this session, you should be able to:

1. Understand the purpose and use of needs assessments, situation analysis, baselines, evaluations and real-time reviews
2. Define what a baseline is and why it is a required activity for monitoring and evaluation
3. Describe the key steps in planning and undertaking baselines and evaluations
4. Describe key differences with needs, assessments, reviews, etc.

Practical and Professional Skills

When you have studied this session, you should be able to:

1. Describe key steps in planning and conducting a baseline survey
2. Select an appropriate evaluation design (to use with different types of interventions)
3. Plan, manage and provide technical inputs into the design and process of a baseline, needs assessment, evaluation or real-time-review, including:
   - Design terms of reference
   - Identify appropriate consultants
   - Use and disseminate evaluation learning and findings for accountability, decision-making and programme quality improvement

1 The purpose and use of baselines and evaluations

1.1 Introduction

Activity 1 (SAQ)

Prior to discussing some definitions, we will investigate some scenarios. Fill in the blank spaces overleaf using one of the following terms:

- baseline;
- (child rights) situation analysis;
- evaluation of humanitarian action;
- operational research;
- needs assessment / initial rapid assessment.

You should then refer to the Answers to the Self-Assessment Activities at the end of this module.

Scenario 1: In a country office, the context has changed over the past year. Therefore, Save the Children decides to do a _________ to determine what are the causes of the displacement; what are its effects on different groups; how do different people perceive it; what is the political/duty bearers context; what different groups and institutions are stakeholders; what are the links between different organisations; what policies are concerned and how are they implemented.

Scenario 2: Monsoon rains have severely affected a programme area. Save the Children decides to carry out a _________ to find out more about what has happened, and which children and how many of them are affected.

Scenario 3: A new 3-year project is starting and ~~as~~ one of the first activities of the project's implementation is carrying out a _________ to establish what are the values of project outcome indicators in this first year. This information will then be used to set targets against which future change in the target population will be measured.

Scenario 4: Ten months into the emergency response that was carried out after the monsoon rains disaster described in scenario 2, Save the Children plans to carry out a _________ in order to assess the extent to which the response has met its objectives, the technical strength of the programme. Save the Children also plans to measure the extent to which the response has been accountable for the affected population and recommend improvements for longer-term strategies.

Scenario 5: As part of _________, a randomised control trial is conducted within an innovative health programme searching for knowledge on strategies, interventions, or technologies that can improve its results. The focus is on improving the health interventions within the programme implementation period and on helping programme managers better understand constraints and new opportunities so that decisions can be based on evidence.

Now I will discuss what the differences between baselines and evaluations are …

1.2 What is a baseline?

Baselines are data collected at the outset of a project (or an activity) to establish the pre-project conditions against which future changes amongst a target population can be measured.

Referring to the Save the Children Baselines Essential Standard, it says that: 'Projects and programs establish a baseline (or other appropriate equivalent) as a comparison and planning base for monitoring and evaluations'.

The information gathered (and analysed) in the baseline consists of data on indicators specifically chosen to monitor project performance on a regular basis. It also considers the anticipated use of these indicators at a later time to investigate project effects and impacts. Indicators (and methods of data collection and analysis) used in baseline surveys may be qualitative or quantitative. Baseline studies can also serve to confirm the initial set of indicators to ensure those indicators are the most appropriate to measure achieved project results.

Activity 2 (exploratory)

Take a moment to think about different ways to get this type of information. How do you collect baseline information in your programmes? What methods do you use?

Did you include approaches other than different surveys? Looking back at the Baseline Essential Standard above, you can see that it is not necessarily saying that you need to carry out a baseline survey. Indeed, the process by which baseline data is collected can be flexible, but establishing these pre-programming conditions needs to be done. For example, this could be carried out by a survey or an appropriate equivalent activity. One such equivalent activity in emergency contexts is given below.

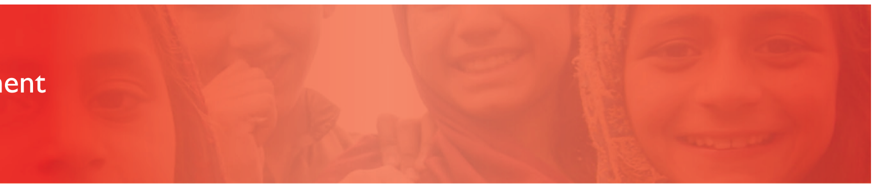
In emergency responses, available baseline data in the Emergency Preparedness Plan may be complemented with rapid assessments. If a baseline cannot be established while prioritising delivery of a timely response, then a needs assessment can be carried out. For example, in a sudden onset emergency, initial rapid assessments are undertaken within 24-72 hours and followed by in-depth multi-sector assessments. While these are different from a baseline study, they can yield the same type of information on pre-programming conditions in emergency contexts.

It is worth noting that baseline studies are different from (child rights) situational analyses which are undertaken to inform the design of a project/program and are more open ended. Indeed, situation analysis is a process designed to collect and analyse information about a particular issue in its broader context in order to make a sensible assessment of what needs to be done. Baseline studies, however, aim to measure the initial status of *specific* project objectives and indicators, and provide the basis for comparison over time. These are strongly linked to a project or program M&E plan.

Some differences between a baseline and a situational analysis:

Baselines are:

- ✓ more specific
- ✓ carried out during the project/program
- ✓ more precise
- ✓ often collect primary data
- ✓ used for setting targets



1.3 What is evaluation?

There are many definitions of evaluation: most of them agree that it is ‘a systematic assessment of an on-going or completed project, programme or policy, its design, implementation and results’.

Evaluation also refers to the process of determining the worth or significance of an activity, policy or program. Evaluations should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision–making process of both recipients and donors.

As it is layed out in the table below, evaluation is different from monitoring. Monitoring can be defined as ‘a continuing assessment based on systematic collection of data on specific indicators as well as wider information on the implementation of projects’. Monitoring enables managers and other stakeholders to measure progress in terms of achieving the stated objectives and the use of allocated funds.

Monitoring	Evaluation
… is ongoing	… is periodical
Gathers information related to the programme regularly, on a day-to-day basis	Assesses the programme’s design, processes and results
Provides information on whether and how the planned activities are implemented or not	Provides information on what the programme effects are
Refers to activities, outputs and intermediate results/outcomes	Refers to intermediate results/outcomes, and the bigger/strategic objectives and Goal (in case of impact study)
Fosters informed re-design of project methodology	Can foster informed review of both the current project (mid-term evaluations) or new projects (final evaluations)
Collects data on project indicators	Uses data on project indicators collected through the monitoring process
Performed by project staff (Internal)	Can be performed by an external evaluator, but also by a mixed team with internal and external evaluators or even by only an internal team.

Table 1: Differences between monitoring and evaluation. Source: adapted from Save the Children M&E training modules

There are many types of evaluations and also a series of related terms, including: proof of concept evaluation, post-programme evaluation, thematic evaluations, real-time evaluation, review, ex-ante evaluation, ex-post evaluation, formative evaluation, summative evaluation, mid-term evaluation, process evaluation, etc. These are described in more detail in the Evaluation Handbook referenced at the end of this module.

Particularly worth mentioning here, however, are Real Time Reviews (RTRs) and Evaluation of Humanitarian Action (EHA) in emergency contexts. RTRs combine two previously separate exercises: Real Time Evaluations and Operational Reviews. Therefore, an RTR looks at both programmatic and operational aspects of a humanitarian response. The primary purpose of an RTR is to provide feedback in a participatory way in real time (i.e. during the implementation of an emergency response, and during the review itself) to those executing and managing a humanitarian response: this is designed to improve operational and programmatic decision making. RTRs are internal exercises that aim to have a light touch so as not to interfere with programming.

Save the Children conducts Real Time Reviews in all Category 1 emergency responses. The decision to have an RTR in Category 2 emergency responses will be based upon discussions with the CO. After Action Reviews are recommended for Category 3 and 4 responses. RTRs take place within three months of categorisation.

RTRs examine the initial phase of the response. The findings and recommendations should help staff to make immediate adjustments to both programming and operations in order to better meet the needs of beneficiaries.

An Evaluation of Humanitarian Action is a systematic and impartial examination of humanitarian action intended to provide lessons to improve policy and practice and enhance accountability in emergency responses. It is just one type of evaluation with distinct characteristics looking at the OECD DAC criteria but also, for example, our Emergency benchmarks, and the context of an emergency that can make access to data more difficult, etc.

See also session 18, ‘Measuring results in Humanitarian work’ for more details on RTRs and EHA.

1 Steps to conduct baselines and evaluations

Activity 3 (SAQ)

The box below has some key steps that would need to be undertaken for conducting an evaluation. On a piece of paper, list these in a correct order. Then also reflect on any steps that may be missing.

Are there any steps that are different for a baseline study?

Involving stakeholders	Designing and testing the evaluation tools	Preparing a management response to the report
Data management	Agreeing a timeline	Sharing findings with stakeholders
Putting the evaluation team together	Training	Reviewing the evaluation report
Making an evaluation plan	Developing an action plan	Approving the evaluation report
	Setting a budget	Drawing up ToR
	Writing an evaluation report	

2.1 The study process

The chart below describes the key steps in a baseline or evaluation study process: from designing the study, through planning and implementing it, to using the results.

In the design phase, we clarify the purpose of the survey, define the objectives and questions, and decide on the design and methodology. It is here that you need to consider issues like the level of rigour, data sources, sampling and data collection, and data quality. In section 3 below I have presented different evaluation models and designs.

The planning and implementation processes include discussing how stakeholders will be involved, preparing a budget and timeline, and forming the survey team. It is also here that you should think about the management process, which includes defining roles and responsibilities for the evaluation team, and preparing the terms of reference and contracts. In the inception phase, you would then carry out more detailed planning and test the data collection tools and other methods of measurement . involving children and stakeholders.

During the implementation of the study, in addition to the data collection, input, cleaning, analysis and storage, writing of the evaluation report should also start.

Finally, it is important to take time and use resources to communicate and disseminate the findings, plan follow up(s) on any recommendations and prepare the response to the evaluation and the management action plan. This should also include ways to ensure that learning informs future project or programme planning, design and implementation. Session 9 'Use of MEAL data' explores in more detail how to use the evaluation results.

The process represented in the graph below can be adapted to different types of evaluations, regardless of the context and the methodology chosen. This process can also be helpful when planning baselines or end-line studies or different types of research, although as indicated earlier in this session, these are different from evaluations.

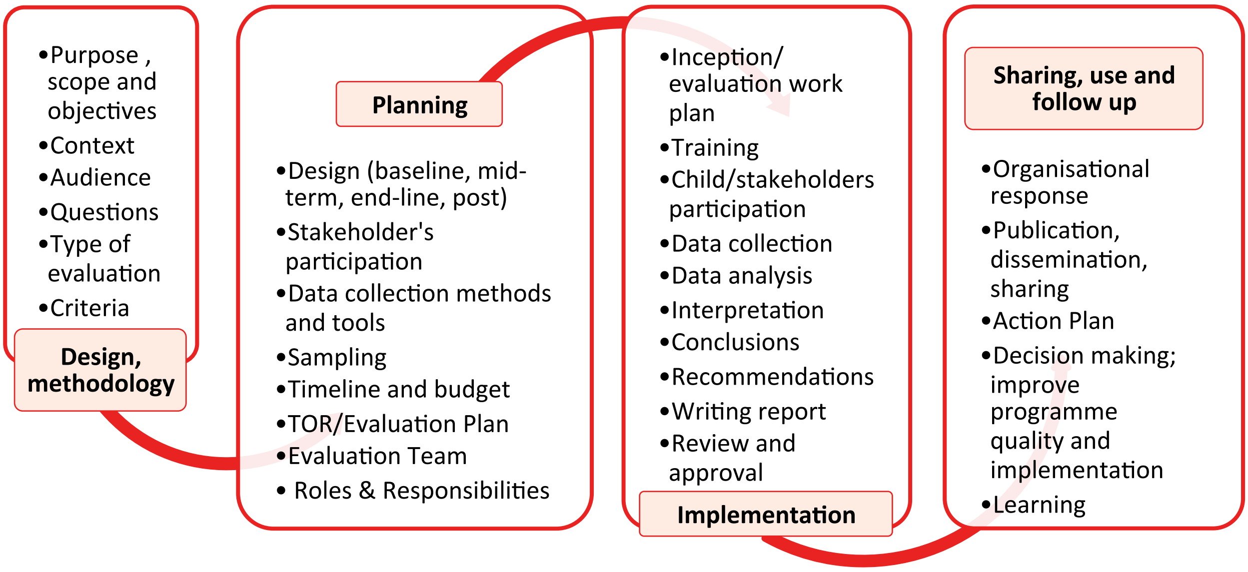


Table 2: Evaluation processSource: Adapted from SC Evaluation Handbook and corresponding M&E Training modules

You may want to have a break here before we explore different evaluation designs!

2.2 Choosing the evaluation approach and design

Evaluation models and approaches

In recent years conducting evaluations has become a frequent practice in the international cooperation sector. Donors demand results and value for money, as well as transparency and efficiency in the use of resources. There is a shared concern for measuring interventions to see if what we are doing works, and to what extent. Evaluations are used for accountability and decision making, but also for improvement and learning, or *enlightenment* for future actions, in the words of Stufflebeam and Shinkfield (1987:23).

They try to not only answer questions such as 'Has the intervention achieved its objectives?', but also, 'How were the objectives achieved?' It is important to identify why you want to conduct an evaluation and what will be the key questions you want the evaluation to answer, so you can choose the most appropriate approach and design. For instance, if you want to conduct an evaluation primarily to show results and to be accountable for the use of resources, you may want to opt for a '**criteria-based standardised model**'. This approach is widely used in the international development sector and it fits very well with objectives and results-based planning tools such as the log-frame. This evaluation approach uses pre-established criteria to organise the evaluation questions and the evaluation design, such as the DAC criteria[1]: effectiveness, efficiency, relevance, impact and sustainability. The criteria-based model assesses the programme's performance against some pre-established standards that determine if the objective has been fulfilled or not.

However, there are other approaches that go beyond the fulfillment of objectives and focus on other aspects of the intervention. For instance, **Theory-based evaluation approaches** (initially proposed by Carol Weiss and adapted by other authors) allow a wider look at the programme. This approach considers the intervention as a system with inter-related components that need to be looked at during the evaluation to understand how the changes are produced. It will help you to read ~~reading~~ the information relating to the programme's 'black box'. This is particularly useful if you plan to use your evaluation for learning and decision making or if you want to document your programme for it to be replicated or escalated.

Example

A sick patient visits a doctor seeking a cure, but the symptoms don't improve after the treatment. Can we conclude directly that the doctor is not a good doctor? Or that the patient is not a good patient? Or that the treatment is not a good one? And what if it has worked for other patients?

There are multiple reasons why results are achieved or not. If the evaluation focuses only on measuring results and the achievement of objectives it will only provide us with part of the picture. We would still not know *why* and *how* the results have been produced and *what is needed* to improve it.

In advocacy interventions, where the use of experimental or quasi-experimental methods is not feasible, you would utilise specific approaches, such as outcome mapping, contribution analysis and process tracing. If you are interested on learning more about evaluation in advocacy, please see session 10, 'Tracking of advocacy'.

[1] These five criteria were proposed by the Development Assistance Committee (DAC) and are now broadly used in the development sector. The criteria were first laid out in 1991 in the DAC Principles for Evaluation of Development Assistance and later defined in 2000 in the Glossary of Key Terms in Evaluation and Results Based Management.

Outcome Mapping considers that development is essentially about people relating to each other and their environment, so it focuses on people and organisations. It shifts away from assessing the products of a programme to concentrate on changes in behaviours, relationships, and also the actions and activities of the people and organisations with whom a development program works directly.

Contribution analysis compares an intervention's theory of change against the evidence in order to come to conclusions about the contribution that it has made to observed outcomes. It examines to what extent observed results are due to the activities of an advocacy initiative rather than other factors, and whether the advocacy initiative has made a difference, and whether or not it has added value.

Process tracing is a qualitative research method that attempts to identify the causal processes – the causal chain and causal mechanism – between a potential cause or causes (e.g. an intervention) and an effect or outcome (e.g. changes in local government practice). The method involves evidencing the specific ways in which a particular cause produced (or contributed to producing) a particular effect. A key step in process tracing is considering alternative explanations for how the observed outcomes came about, and assessing what evidence exists to support these.

Rights-based approaches organise the evaluation questions and design around the rights that the intervention intends to promote, protect or fulfill. The United Nations Convention of the Rights of the Child (1989) is the primary Human Rights framework for child-rights based programming and evaluation. The rights and principles presented in the convention are usually organised around four groups: survival and development; non-discrimination; participation; and the best interest of the child. This approach puts children at the centre and recognises them as rights holders. It analyses power relations between groups, in particular between duty bearers and rights holders, focusing on the most vulnerable groups with a particular interest on how principles such as non-discrimination and participation have been integrated into the programme. It prioritises sustainable results, looking at the real root-causes of the problems instead of just considering their immediate causes.

Which one is the best approach?

It all depends on the purpose of your evaluation, what you expect the evaluation to tell you and how you and other stakeholders are planning to use the evaluation. You also need to check if the donor is requesting a particular approach or if it is possible to be propositional about it. These models can also be complementary, so you can combine elements of several of them to create a tailored model. For instance, you can organise the evaluation questions by groups of rights and sub-grouping them as well by criteria and standards, including questions that relate to the process and the structure and not only to the results.

It is not possible to be completely *objective* when evaluating programmes, so if you want your findings to be reliable, it is important to be as systematic as possible. Here are some tips that can help you:

- Reflect stakeholders' information needs in your evaluation questions.
- Assure the quality of your data collection and analysis techniques
- Follow four separate steps of a systematic assessment: data analysis; data interpretation; conclusions; and recommendations.
- Ensure the highest standards on ethics, stakeholder participation and transparency

2.3 Evaluation design

If your evaluation will measure results, then once you have decided the evaluation approach and the main evaluation questions, you will also need to choose the appropriate evaluation design. This decision will also depend on elements such as the available resources (some designs are more resource-intensive than others) and the implementation stage of the programme. Some evaluation designs require conducting measurements before the programme activities actually start (baselines) so it will not be possible to use them if you only start planning for your evaluation when the programme is in an advanced implementation stage.

Why do you need a good evaluation design?

When we conduct an evaluation we want to find out if there have been any changes (*gross effect),* but also if the changes are produced by the programme. This means: can the changes be *attributed* to the programme? If yes, *how much of these changes* are actually due to the programme activities (*net effect*). This can only be established if we have a sound evaluation design.

Robust evaluation designs present high degrees of:

- Internal validity => It refers to the extent to which all possible explanations or hypotheses are controlled.
- External validity => It refers to the extent to which results can be generalised beyond the object of study and if the same conclusions can thus be applied to other subjects, places or populations. This depends on the quality of the sampling, the margin of error we are dealing with and the control of potential bias.

You can find further information on these concepts in session 6 –'Methods of data collection and analysis' in this training module.

There are two strategies to ensure the internal validity of your evaluation design. These can also be combined. You can (1) compare groups between themselves or over time or you can (2) use statistical analysis.

1. **When you compare groups between themselves and/or over time** you can see their differences and eliminate the effect of possible explicative alternatives you may be considering to explain the changes.

   In order to do this, you can:

   a) Compare the treatment group with the comparison or control groups (not receiving project activities/services)

   b) Compare programme group with itself over time

2. **When using statistical analysis** it is possible to control or maintain constant the effect of possible alternatives. This requires identifying the variables to be controlled and gather data on them. Then use statistical analysis techniques to separate the effects of each variable.

What is a treatment group? (TG)
The group participating in the programme, receiving the services or goods.

What is a comparison group? (N-ECG)
This is a group with very similar (almost identical) characteristics to the treatment group but that is not participating in the programme. It is also called a non-equivalent group because it is almost identical, but the members are not selected with random sampling — thus, possible bias does not make them totally equivalent.

What is a control group? (ECG)
This group is a comparison group – so no participating in the programme – but its members have been randomly selected. This means the group can be considered completely equivalent to the treatment group.

In this session we will concentrate on the first methodological strategy (comparing groups between themselves and/or over time). Let's see what the most known quantitative evaluation designs are to compare groups!

From least to most robust.[2]

Design	How robust? Can calculate net effect?
1. Post-test (only TG) In this model you take one measurement at the end of the intervention with the treatment group. For instance, you can conduct a survey with programme participants at the end of the activities. As there is only one measurement, we cannot compare the data with other groups or with the group itself before the activities, so it is not possible to establish if there have been changes and if these are attributable to the programme. We can only document changes perceived by participants to which the programme will have contributed.	Non-experimental designs. Little internal validity Only possible to extract **gross effect** – the results that the programme is apparently producing or as perceived by those participating in the programme. It includes variables related to the programme as well as other factors, so we can only talk about contribution, not attribution.
2. Pre and post-test (only TG) This model compares the results before and after the programme by measuring changes only in the treatment group. It is not possible to know if there have been any other variables affecting the intervention and the changes, if any.	
3. Only post-test w N-ECG The measurement only takes place at the end of the programme, but both the treatment and the comparison group (=non-equivalent group) are measured. You can compare the results between both groups, but there is no comparable data of the initial situation, before the programme started.	
4. Pre and post-test with N-ECG With this model you will perform measurements before and after the programme to both the treatment group and the comparison group. It is very close to the real experiment, but it is not as robust because the comparison group is not selected randomly. This is why it is called non-equivalent group.	Quasi-experimental design. Medium internal validity.
5. Time series (longitudinal analysis) Here several measurements are taken before the programme starts, while it is being implemented and after it has finished, so you can have a broader picture of how the treatment group is changing, if this is the case. The differences between measurements can be treated as net effect, but we cannot be as sure as with the experimental design. The comparison group (N-ECG) can be included or not included.	
6. Experiment with pre and post-test This model is also called 'Randomized Controlled Trial' (RCT) and it is considered the only real experiment. Here the individuals of both the treatment group and the control group are selected randomly from the same group of individuals. The treatment group receives the programme that is being tested, while the control group receives an alternative treatment, a dummy treatment (placebo) or no treatment at all. The groups are followed up to see how effective the experimental treatment was and any difference in response between the groups is assessed statistically.	Experimental design. High internal validity. Calculates net effect.

Table 3: Examples of quantitative evaluation designs. Source: self-elaborated

[2] This list is not exhaustive. There are other models available, but these represent a good menu to choose from when planning an evaluation. Further information on other quantitative evaluation models are presented on the book 'Real World Evaluation', referenced at the end of this session.

The **non-experimental designs** (models 1, 2 and 3) are not very robust and only provide the gross effect of a programme, but require fewer resources and the analysis is not complex. Sometimes, if the programme has already started, models 1 and 3 are the only ones that are available for us to do. For this reason it is worth planning the evaluation during the programme planning stage, so a pre-test can be planned for and performed before the activities start.

The **experimental design** (number 5) provides the soundest methodology to extract the net effect, but it is expensive and can very often pose ethical problems, because individuals may be excluded from a programme just so they can be treated as the control group. For this reason Campbell introduced quasi-experimental models (such as model 4 and 5) that provide an acceptable degree of internal validity while posing less ethical issues.

Strictly speaking, impact evaluations are those which provide the net effect: this means that they calculate the 'amount of change' that can be claimed by the programme (**attribution**). Otherwise, we can only say that the programme has **contributed** to the changes.

Let's see some practical examples:

In the example below, if we only look at the treatment group (blue line) there seems to be a 20 points gain (45-25=20). However, a closer look will show you that the comparison group, who has not received the programme, has gained 10 points (35-25=10). This means that the programme has only produced 10 points gain. The other 10 may be due to other factors, but we can only see this if we compare the results between to two groups.

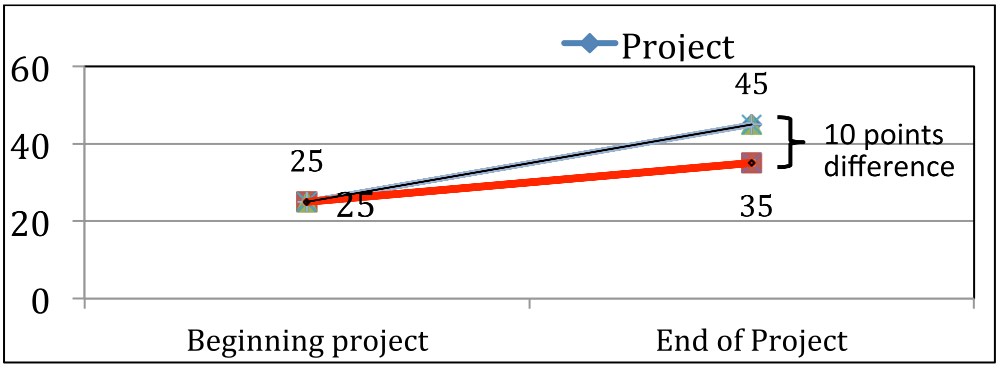


Figure 1: Difference between baseline and endline (example 1). Source: elaborated by Larry Dershem

In the second example no real change can be attributed to the programme because even if there is a gain of 20 points, the same gain has been recorded in the group who is not receiving the programme. Thus, changes are due to other variables or correspond to a natural trend.

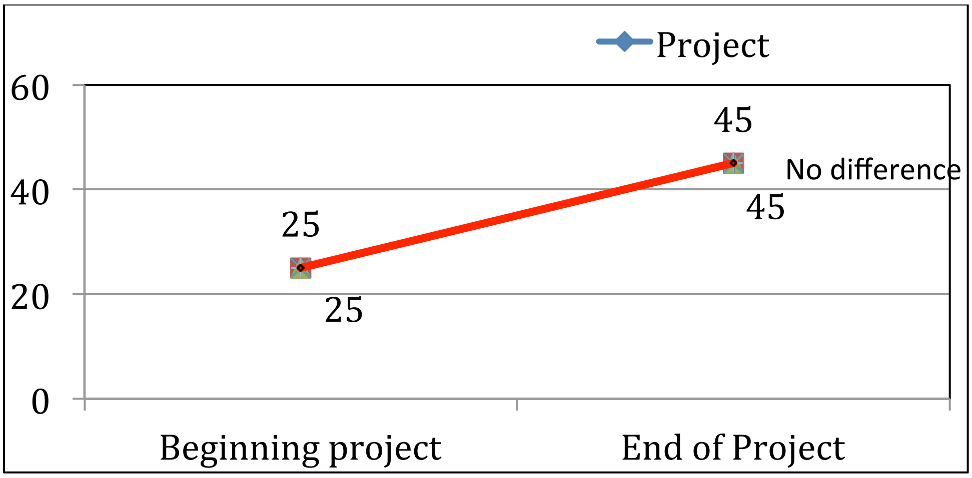


Figure 2: Difference between baseline and endline (example 2). Source: elaborated by Larry Dershem

The same happens in the third example, even though both groups start at different points, they both make the same increase (45-25=20 for the project group; 35-25=20 for the non- project group).

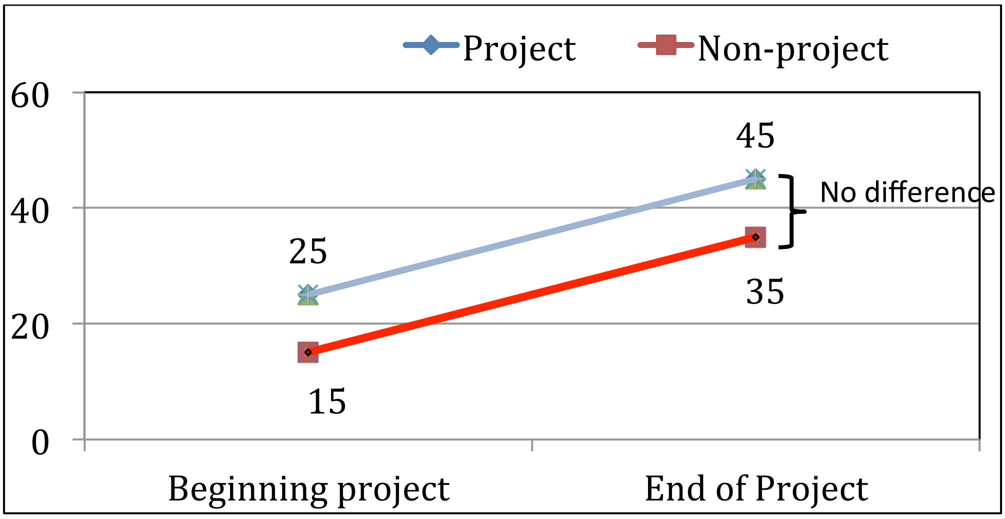


Figure 3: Difference between baseline and endline (example 3). Source: elaborated by Larry Dershem

Activity 4 (exploratory)

Read the following scenario and propose one or two evaluation designs, as appropriate. You should also think of the advantages and limitations of each of them.

Pneumonia is one of the world's leading causes of morbidity and mortality in children, causing roughly 1.6 million deaths per year. In country X Save the Children is working with a national-wide network of community based health workers (CBHW) who provide preventive and basic curative healthcare to women and children <5 years of age. In routine, these CBHWs manage simple pneumonia with cotrimoxazole for 5 day – whereas they refer the severe pneumonia cases to the nearest health facility with limited referral compliance. In order to make the management of severe pneumonia easily accessible to the communities, the CBHWs were trained in identification and management of severe pneumonia with oral amoxicillin with excellent results. In order to try and secure more funding, it was decided that there was a need to evaluate the pilot project and present evidence of its impact and scalability.

(Source: Extracted from Pakistan's application to the 2012 Members Meeting Program Innovation Competition. Full document available in OneNet. Click link here. The full study was published by the journal The Lancet and is available here: http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(11)61714-5/fulltext).

To summarise:

- There is a variety of programme evaluation approaches to choose from. The most widely-known and used in this sector is the criteria-based standardised model. But you can also consider other options and methodologies, such as theory-based models, outcome mapping, process tracing, contribution analysis and rights-based approaches if they are more adequate for your intervention and context.
- Evaluations designs will determine if the evaluation is able to provide the gross effect (changes that occurred) or the net effect (changes that can be attributed to the programme).
- In this session we have explored the main quantitative evaluation designs and discussed their degree of internal validity. From less to more robust we have seen: post-test only, pre-post-test, only post-test with non-equivalent control group, pre and post-test with non-equivalent control group, time series, and experiment with pre- and post-test.

3 Planning and managing baselines and evaluations

3.1 Introduction

As mentioned in the section above on evaluation process, conducting evaluations is a process that starts with a discussion of the motivations for conducting the evaluation. Additionally, the evaluation should only be considered finalised when the results are disseminated, the evaluation findings are used and the recommendations widely discussed.

Consulting with a wide range of stakeholders (and reaching agreements) can take quite some time and significant preparation, particularly when involving children. If you need to plan for an evaluation, make sure you allow enough time for each step. A full evaluation process may take several months (though usually less in real-time evaluations or evaluations of humanitarian interventions).

3.2 Who should participate in the evaluation process? Identifying and engaging stakeholders

In order to apply the principles of participation, transparency and accountability, your evaluation process should clearly define a management structure and stakeholder input, laying out who is going to do what.

Activity 5 (exploratory)

Reflect for a moment on your projects/programmes: who would be on the evaluation team, who would manage the evaluation process, and who are the groups of stakeholders that would be involved in an evaluation? Then compare with the list below.

Roles and responsibilities of the key players in an evaluation process should be set out in the terms of reference (ToR) and the M&E plan. More details on the ToR are given in the corresponding section below. Key people involved in an evaluation would include:

- ✓ Thematic, programme or project manager(s), who'll have the overall responsibility for managing the evaluation process. It is worth noting that it is important that (s)he is able to keep sufficient distance and in case there is a conflict of interest, that this is declared and discussed before the evaluation.
- ✓ M&E and MEAL staff: one of these team members could be the evaluation manager, and they could support the evaluation with other technical/methodological input or assistance.

Activity 5 (exploratory) ***continued***

- ✓ Country Director, Director of Programme Development and Quality (PDQ) and the Senior Management Team (SMT) keep an overview of the evaluation, and participate in key meetings (e.g. debrief meetings) with the evaluation team, take part of any reference groups[3] set up, and are accountable for the implementation of action plans arising from the evaluation process.
- ✓ Advisors at different Global Initiative, Regional Office, Member or Head Office levels, including MEAL advisors for emergency programmes, should be available to input in the ToR, evaluation plan, inception report, selection of consultant(s), etc.
- ✓ In terms of external stakeholders playing a role in the evaluation process – children, communities, donors, government, academic and civil society organisations should be considered to be part of the evaluation team, or at least part of the reference group or other feedback mechanism.

3.3 The evaluation manager

All evaluations should appoint an evaluation manager. This person will assume day-to-day responsibility for managing the evaluation process. Their role is distinct from that of the evaluation team leader, who will lead on content and implementation. The evaluation manager is also responsible for logistics and for facilitating the broader evaluation process, including quality control. The evaluation manager's responsibilities are centered around deliverables in the evaluation plan (such as the stakeholder analysis, the evaluation questions, the field work plan, finalisation of data collection tools, management of data, report dissemination, etc.).

The evaluation manager also ensures that a quality assurance takes place. This will include looking at

- ✓ the credibility of the findings
- ✓ how well the work addresses the aims of the evaluation
- ✓ the rationale and robustness of the evaluation design
- ✓ the way the sample was taken
- ✓ the data collection processes
- ✓ comparing the draft report with the ToR

An example checklist for quality review is annexed to Save the Children's Evaluation Handbook (Annex 4).

---

[3] You may want to set up an advisory group or reference group that involves key stakeholders, including children and young people. A reference group can make the evaluation process more accountable and can help to ensure buy-in from stakeholders, including donors. The reference group(s) might, for instance, input to the ToR and comment on your plans for the evaluation questions, design and methodology. If you are evaluating an emergency response programme, it may be particularly useful to set up reference groups, especially when there are several Save the Children members with a large stake in the programme.

3.4 Preparing the Terms of Reference (TOR)

During the planning phase of your evaluation, baseline or real-time review, you will need to prepare the Terms of Reference (ToR). This is a structured document that describes the reasons for conducting the study (purpose), its objectives and its scope, amongst other things. It should cover what we want to evaluate, why, and how it should be organised.

As indicated in the Save the Children Evaluation Handbook, a ToR can also serve as a management tool throughout the evaluation process and a reference against which the completed evaluation report and deliverables are assessed before approval. The scope and level of detail of a ToR might vary; in some cases, the ToR will entail a plan for the entire evaluation process, while in others, ToRs can be developed for specific parts of the evaluation – e.g. for the evaluation team or even individual team members. Furthermore, the scope and level of detail in the ToR should be consistent with the size, complexity and strategic importance of the project or programme being evaluated.

As you will have already guessed, stakeholders should be widely consulted while the ToR is being developed, so their views are clearly reflected and their information needs included. It will also help to clarify expectations around the process and to motivate stakeholders to be fully involved. The success of the evaluation and its use depends on it!

Very often the donor will provide a ToR template that should be followed. If this is not the case and you need to develop your own, it can be useful to collect a few examples for inspiration.[4]

Activity 6 (exploratory)

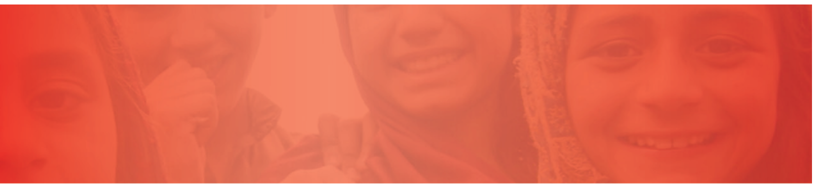
What information should the terms of reference (ToR) contain?

Think of a particular project you are working on or one that you are familiar with and write down the main headings of the ToR that you would suggest for its evaluation (up to 10).

Now compare with the ones I came up with. These are some headings you will often find in evaluation ToRs, whichcan also be adapted for baselines, end-lines or other type of reviews.

- **Background** => Description of the project/programme being evaluated, including objectives, timeframe, target areas, beneficiaries, relevant financial data, etc. Anything important which will help to explain the context in which the evaluation will be conducted should also be mentioned.

[4] Save the Children International does not have an official template for evaluations ToR, but you can find a good example in the Save the Children Evaluation Handbook (Annex 3).



Activity 6 (exploratory) *continued*

- **Purpose of the evaluation** => This section should clearly describe *why* the evaluation is being conducted.
- **Objectives and key evaluation questions** => These should address the information needs identified during the initial phase and will influence the methodological choices that need to be made later on. The timeline and available resources should be taken into account when defining the objectives – it is important to be realistic and plan well for what can be achieved. The evaluation questions will be extracted from the main information needs that have emerged during the first rounds of consultation, but once the process as such starts, additional evaluation questions may be included.
- **Scope of the evaluation** => related to the previous point, this section should define what the evaluation will cover in terms of geographic areas, time-frame and programmatic aspects.
- **Evaluation design and methodology** => Some ToR already present the approach, design and methodology to be followed and it can be very useful if this has been discussed and agreed by all stakeholders, including the evaluation team. But some ToR do not include this, as managers decide to discuss the methodological aspects with the evaluation team and allow them to make a technical proposal. In this case, the methodology can be described in an inception report, a document that describes in more detail how the evaluation will be conducted.
- **Organisation, roles and responsibilities** of the different groups of stakeholders
- **Outputs or expected deliverables,** such as – *Preparation and desk review phase*: inception report, list of documents; work plan; evaluation matrix, list of stakeholders; field work agenda; data collection tools. *Field work phase*: field work report; raw data. *Analysis*: draft report; final report; executive summary. *Communication phase*: dissemination plan and tools.
- A **timeline** describing when the different steps should take place
- A list of the **budget and resources** needed to conduct each activity within the evaluation process
- A **plan for disseminating** the findings and sharing learning

3.5 Recruiting and managing the evaluation team

A key factor in the success of the evaluation process is the evaluation team. When planning the evaluation you will need to decide if it is more appropriate to have a team of internal evaluators, external evaluators or a mix of both. External evaluators are usually less influenced by the institutional context but may not be familiar with the details of relevant organisational issues affecting the project. External evaluators are usually perceived as more independent, although internal evaluators should not be directly related to the intervention either. There is not a right or wrong answer for this, so you may want to discuss with the reference groups what is the best arrangement for your specific evaluation.

What skills should you ask for? The ideal case is to have an expert to evaluate each thematic area addressed by the intervention, and also an expert on evaluation and methods. For our programmes, it is also particularly important that the team has the skills to work with children and involve them in the process as much as possible, although orientation and training may be needed for this. Evaluators should speak the local language and know the country context as much as possible, so local evaluators are usually preferred, although international evaluators could also provide additional capacity to a local team.

Evaluating is not only collecting and analysing data. As noted by C. Weiss, it is a process that happens within a political context. There are a number of interests around the evaluation results that will influence the process and evaluators should be sensitive to this reality.

If the team is composed by several evaluators it is best to name a team leader. It usually works better when members have worked together before.

When looking for external evaluators, it is considered good practice for transparency to advertise your evaluation in different networks and platforms. This allows all potential candidates to submit an Expression of Interest (EoI) or a proposal and participate in a competitive process. This also increases your chances of getting the best value for money.

Activity 7 (exploratory)

Think of several different ways to advertise your evaluation to find an evaluation team?

There are different ones you may have thought of in your context, but here there are some I can suggest, although not all of them are free of cost: country office website; evaluators rosters, if available; professional and personal networks (internal and external); external websites related to development or evaluation (Relief web; Devex, evaluation societies…). You can also inform your relevant Regional Office, members and Global Initiatives so they help you disseminating your ToR.

Activity 7 (exploratory) *continued*

Once you have identified potential candidates based on their skills, experience and qualifications, it is a good idea to arrange an interview with shortlisted candidates to make up your mind. Whenever possible, ask members of the reference group, your M&E advisor or project managers to take part in the interviews, for transparency. It is a good practice to have an assessment tool to rate the candidates proposals, CVs and responses during the interviews.

Once you have a good team you will need to negotiate and agree a contract. Remember that they all have to subscribe to our child safeguarding policy! Allow enough time for this, as it is important that all expectations, roles and responsibilities are clear from the beginning. This will save a lot of issues and misunderstandings later on. It can be useful to have a manager tool with the work plan and deliverables so you can track progress.

3.6 Sharing and using the results

As you will see in Session 9, 'Use of MEAL data', using evaluation results is part of the way for us to learn what does and does not work. By doing this, we can adapt programming in light of the findings for continual improvement.

Activity 8 (exploratory)

Reflect back at the latest evaluation you have been involved in. How were findings from this evaluation disseminated? Were these shared with children at all? If and how were recommendations used for accountability, decision- making and learning? Then have a look at the suggestions below for sharing and using evaluation results as well as those referred to in the Save the Children Evaluation Handbook.

Communicating and disseminating evaluations with various stakeholders –including children –is a vital part of how we are accountable to them. Referring to the Evaluation Essential Standard and its Procedure, the evaluation should include a plan of how you intend to disseminate and communicate findings. In this plan, you should consider including:[5]

- ✓ a two-way feedback at community, district and/or national level
- ✓ providing an opportunity for beneficiaries and stakeholders to respond
- ✓ telling respondents in the evaluation about their participation and how their contributions have been used
- ✓ sharing with regional office, centre, Global Initiatives and members

[5] The Save the Children Evaluation Handbook has different examples and suggestions for sharing evaluation findings with external audiences and stakeholders.

The management response and action plans are the key ways you manage the use of the evaluation results. The management response to the evaluation lists the key recommendations, along with actions proposed by the evaluation and also the response of the project or programme team. The management response is often part of the evaluation report and hence a public document.

The evaluation action plan is developed for our internal management purposes and is more detailed than the management response. It is a requirement for Save the Children as per the Evaluation Essential Standard and the Country Director is accountable for its follow-up and implementation. In terms of content, it is similar to the MEAL Action trackers described in the Use of MEAL Data module: the action plan includes recommendations and lessons learned during the evaluation, together with actions to address these with resource requirements (financial and/or technical), a timeline and a list of who is accountable for which actions.

Summary of this Session

1. **Concepts and definitions:**
    - Baseline studies are different from a needs assessment or situational analysis
    - Only collect baseline data on your project/program indicators
    - For emergency responses, available baseline data in the Emergency Preparedness Plan should be reviewed and used. In Category 1 or 2 rapid onset emergencies, initial rapid assessments should be undertaken within 24-72 hours
    - Evaluation is the systematic and objective assessment of an on-going or completed project, program or policy – its design, implementation and results.
2. **When deciding the best evaluation approach and design, consider:**
    - The evaluation purpose
    - The evaluation questions
    - Time
    - Resources
    - Implementation stage
    - Data availability
    - Ethical issues
    - Desired level of rigour
3. **Planning and managing baselines and evaluations**
    - Planning baselines and evaluations is not only about preparing for data collection and analysis. It's a broader process that should be participatory. Make sure you allow enough time for it!
    - A good ToR will include all the elements you need to plan and manage the evaluation
    - The higher the involvement and ownership from stakeholders, the higher the chances are that the evaluation and baseline will be used.

Additional resources:

If you would like to learn more, please see:

Save the Children Evaluation Handbook (2011). Also available in ES, FR and Arabic. https://onenet.savethechildren.net/whatwedo/me/Shared%20Documents/Archive%20information/Monitoring%20and%20Evaluation%20Reference%20Materials/STC-evaluation-handbook%20final.pdf

Handbook of Management of Evaluations of the Spanish Cooperation (2010). Available in English: http://www.exteriores.gob.es/Portal/es/SalaDePrensa/Multimedia/Publicaciones/Documents/Cooperacion/Evaluacion/Metodologias/ManualEvaluacion.%20versión%20inglesa.pdf

And Spanish:
http://www.exteriores.gob.es/Portal/es/SalaDePrensa/Multimedia/Publicaciones/Documents/Cooperacion/Evaluacion/Metodologias/Manualdegestióndeevaluaciones.pdf

IFAD's Evaluation Manual – Methodology and processes (April 2009).
http://preval.org/en/content/ifad%C2%B4s-evaluation-manual-2009

Real Team Evaluation Real-time evaluations of humanitarian action (March 2009). An ALNAP Guide. Pilot Version. John Cosgrave, Ben Ramalingam and Tony Beck..
http://www.alnap.org/resources/guides/evaluation/rte.aspx

Integrating Human Rights and Gender Equality in Evaluation. Towards UNEG Guidance. United Nations Evaluation Group (2012). Available in EN, ES, FR and Arabic.
http://www.uneval.org/papersandpubs/documentdetail.jsp?doc_id=980

Evaluation. Methods for studying programmes and policies (1998) Carol H. Weiss. Prentice Hall. Second Edition

Bamberger, M., Rugh, J. and Mabry, L. (2012) *Real World Evaluation. Working Under Budget, Time, Data, and Political Constraints*. Second Edition. ISBN: 9781412979627.
http://www.sagepub.com/books/Book234002

Patton, M. Q. *Utilization-Focused Evaluation* (4th edition). Sage, ISBN 1-4129-5861-X ISBN 978-1-4129-5861-5.

Mayne, J. (2008) '*Contribution analysis: An approach to exploring cause and effect*', ILAC Brief 16, May 2008.
http://www.cgiarilac.org/files/publications/briefs/ILAC_Brief16_Contribution_Analysis.pdf

Oxfam, Process tracing: Draft Protocol
http://policypractice.oxfam.org.uk/blog/2013/02/~/media/C396B507E01C47AB880D7EEF9ECCD171.ashx

Answers to Self-Assessment Question (SAQ) Activities for Session 5

Activity 1, section 1.1 – The purpose and use of baselines and evaluations

Scenario 1: (child rights) situation analysis
Scenario 2: needs assessment / initial rapid assessment
Scenario 3: baseline
Scenario 4: evaluation of humanitarian action (EHA)
Scenario 5: operational research

Activity 2, section 1.2 – Key steps in planning and undertaking baselines and evaluations. To answer our second question, you have probably noticed that all these steps can be relevant for both baselines and evaluations.

1. Draw up ToR
2. Involve stakeholders
3. Setting budget
4. Agreeing timeline
5. Putting the evaluation team together
6. Making evaluation plan
7. Designing and testing the evaluation tools
8. Training
9. Data management
10. Writing evaluation report
11. Reviewing evaluation report
12. Approving evaluation report
13. Preparing a management response to the report
14. Developing an action plan
15. Sharing findings with stakeholders

Activity 3, section 2.1 – Evaluation design and approach

For this programme I would propose a **theory-oriented approach** and would look at how different dimensions of the programme (such as structure and processes) work together to produce the results, so that I can measure results but also understand how this results have been achieved. I would use both **quantitative and qualitative** methods. Quantitative data will measure the effect of the programme on the population in quantifiable manner, and qualitative data will provide us with more insight on the processes, the behaviours and the beliefs that contributed (or did not contribute) to the achievement of the given results. Hence, this would be an evaluation that used mix methods.

As per the **evaluation design**, the team who evaluated this programme opted for a ***cluster randomised trial*** to measure results and extract the net effect. This is a type of experiment – and as such includes pre-test and post-test with the control group.Here, however, groups of subjects (as opposed to individual subjects) are randomised. Some advantages of this design over individual RTC include the ability to study interventions that cannot be directed toward selected individuals and the ability to control for 'contamination' across individuals (i.e. One individual's changing behaviours may influence another individual to change their behaviour). However, the design and analysis involves greater complexity than individual RTC and more participants are required to obtain the same statistical power.

Cover image: