



Practical Guidance on Developing Indicators

MAKING YOUR INDICATORS SMART: SPECIFIC, MEASURABLE,
ACHIEVABLE, RELEVANT AND TIMEBOUND

Author: Velida Dzino-Silajdzic

The case studies on adjusting the donor-mandated indicator (Annex 5) and the Salalem complex indicator (Annex 7) were adapted from facilitation guides for webinar sessions prepared by Dominique Morel and Meghan Krumholtz.

Acknowledgments

To Dominique Morel for her tireless guidance and thoughtful reviews from the programming perspective that significantly enriched the depth and scope of this document. To Clara Hagens, for her challenging yet helpful comments and enthusiastic promotion of the guidance among peers. To the reviewers: Alex Causton, Joseph Junior Sanders, Onugu Emeka, Michael Potts, Marie Miano, Melissa Kreek, Esther Kauffman, Dina El-Araby, Stacey Prieto and Meghan Krumholtz. A heartfelt thank you also goes to over 100 EMECA programming staff who participated in the Essentials of MEAL training in 2019 and 2020, as well as online webinars. Finally, thank you to MEAL advisors for their final comments to ensure relevance of the guidance across CRS offices.

Catholic Relief Services

Catholic Relief Services is the official international humanitarian agency of the Catholic community in the United States. CRS saves, protects and transforms lives in more than 100 countries, without regard to race, religion or nationality. Our relief and development work is accomplished through programs of emergency response, HIV, health, agriculture, education, microfinance and peacebuilding.

Contents

I. What is an Indicator?	1
II. Why are Indicators Important?	1
III. What are the Elements of an Indicator?	2
IV. Quantitative Indicators Should be SMART	3
V. Wording of the Indicator and SMART Dimensions	5
Example 1 from the field: Wording of indicators and questions in data collection tools.....	6
VI. Developing Indicators: Who, When and How?	7
VII. Tools for Development of Indicators	10
VIII. Common Indicator Challenges, Good Practices and Tips	12
Specificity: Typical Challenges and Implications	13
Example 2 from the field: Complex indicators.....	14
Specificity: Good Practices and Tips	15
Example 3 from the field: Improving indicator definitions.....	15
Example 4 from the field: Adjusting donor indicators	21
Measurability: Typical Challenges and Implications	26
Example 5 from the Field: Are you measuring the unmeasurable?.....	27
Measurability: Good Practices and Tips	28
Example 6 from the field: Leading, coincidental and lagging indicators for education project.....	28
Example 7 from the field: Light monitoring in practice	30
Achievability: Typical Challenges and Implications	31
Example 8 from the field: Seasonal effect on attainability	31
Achievability: Good Practices and Tips	32
Relevance: Typical Challenges and Implications	36
Relevance: Good Practices and Tips	36
Example 9 from the field: Education project	38
Time: Typical Challenges and Implications	39
Time: Good Practices and Tips	39
Annex 1: Example—Writing Indicators to include direction of change and target	42
Annex 2: Frequently asked questions about qualitative indicators	43
Annex 3: Using closed-ended questions to measure quantitative indicators	48
Annex 4: Step-By-Step Guidance to Develop Indicators	52
STEP 1: Reflect on the indicators using the draft theory of change	52
STEP 2: Relate the indicators to the types of change / objective statements in the Proframe	54
STEP 3: Prioritize the indicators with their use in mind	58
STEP 4: Refine the indicators and run them through the SMART checklist	60
Annex 5: Adjusting a donor-mandated standard indicator: The case of a BHA indicator	61
Annex 6: PIRS template	68
Annex 7: Salalem complex indicator and how we learned to love it	71
References	78

Abbreviations

BHA	Bureau for Humanitarian Assistance
BPRM	Bureau of Population, Refugees, and Migration
CBE	Community-based education
CRS	Catholic Relief Services
DRL	Bureau of Democracy, Human Rights and Labor Affairs
FGD	Focus group discussion
GAC	Global Affairs Canada
GAIN	Globally accepted indicators
INEE	Inter-agency Network for Education in Emergencies
IDP	Internally displaced person
IPTT	Indicator performance tracking table
IR	Intermediate results
LOA	Life of award
NFI	Nonfood item
NGO	Nongovernmental organization
MEAL	Monitoring, evaluation, accountability and learning
MOE	Ministry of Education
PIRS	Performance Indicator Reference Sheet
PMP	Performance monitoring plan
QLA	Quality Learning Assessment
QUIPS	Qualitative Inquiry Planning Sheets
RF	Results framework
S&S	Shelter & Settlements
SEM	Social Ecological Model
SO	Strategic objectives
SMART	Specific, Measurable, Achievable, Relevant and Timebound
SMILER	Simple Measurement of Indicators for Learning and Evidence-based Reporting
SPICED	Subjective, Participatory, Interpreted, Cross-checked, Empowering and Diverse
TA	Technical advisor
TOC	Theory of change
USAID	United States Agency for International Development



I. What is an Indicator?

Indicators provide evidence that the intended change is occurring.

Indicators are quantitative or qualitative factors or variables that provide a simple and reliable means to measure achievement, reflect the changes connected to an intervention, or help assess the performance of a development actor.¹ Indicators define a measure of change for the objective statements at all levels of the results framework (RF)—strategic objectives (SOs) and intermediate results (IRs)—as well as outputs. When chosen well, indicators measure progress through the project’s pathways of change and enable testing of a project’s theory of change (TOC). While an objective statement identifies what we hope to accomplish, indicators tell us by what standard that result will be measured and provide evidence that the intended change is occurring.²



Indicators support accountability to key project stakeholders.

II. Why are Indicators Important?

Indicators ensure evidence-based decision-making, support a project’s adaptive capacity and advance learning. They help tell the project’s story, thus supporting accountability to key project stakeholders. As such, indicators are at the heart of an effective project’s performance management system or monitoring, evaluation, accountability and learning (MEAL) system.

MEAL system

A MEAL system comprises people, processes, structures and resources that work together as a whole to define, generate, manage and analyze useful and reliable programmatic information for adaptive, results-based project management, evidence-based learning, and reporting and communication to specific audiences.* Indicators are the foundation of an effective MEAL system as they govern what will be measured, how, when and by whom. Their definitions, disaggregation, data collection details and plans for use guide how the analysis should be performed to generate reliable data for participatory interpretation to support adaptive, results-based project management and evidence-based learning.

* Adapted from Glossary of MEAL Terms (CRS 2014)

To be able to use indicators for these purposes, it is necessary to not only measure the indicator, but also identify beforehand a target for that indicator, and very often identify a starting point or baseline. The baseline is the value of a performance indicator before the project starts, while the target is the specific, planned level of result to be achieved within an explicit timeframe.³ Without knowing where an intervention started and what it intends to achieve, it is impossible to properly assess its progress.⁴

1. CRS 2015a.

2. USAID 2010c.

3. USAID 2010a.

4. Global Affairs Canada 2016b.

III. What are the Elements of an Indicator?



Indicators consist of information that signals a change.

In the simplest terms, indicators consist of information that signals a change.⁵ The way that information is presented or the structure of the indicator differs depending on whether the indicator is quantitative or qualitative.

Quantitative indicators produce numerical values as they measure amount or quantity. Examples include: *the number of students that enrolled in Grade 1; the dollar value spent on students' books; the percentage of trainees that passed the knowledge test; and the ratio of women to men in decision-making positions of government.* A quantitative indicator is typically composed of a:

- *Unit of measure*, which could be a number, percentage, ratio or rate.
- *Subject of measure*, e.g., households, project participants, women, kits, vouchers.
- *Description of what is being measured*, e.g., completion of training, reporting confidence to make own decisions, receiving winterization nonfood items before the onset of winter.
- *Disaggregation requirements*, e.g., gender, age, location, vulnerability categories, education level, sector.

Disaggregation requirements can either be part of the indicator wording (e.g., *percentage of refugees reporting that the temporary shelter is appropriate to their needs, disaggregated by gender, location and household size*) or presented in the supporting MEAL design documents (see [Section VI](#)). Either way, relevant disaggregation categories must be specified to meet the information needs of the programming team.

Often, quantitative indicators are worded in a neutral manner, i.e., they don't indicate a direction of change (increase, decrease, improvement), nor embed a target (e.g., 250, 55%). However, some donors require that indicators are written in a way that incorporates these elements (see [Annex 1](#) for an example to help meet these donor requirements).

Qualitative indicators generate narrative information (i.e., text), rather than numbers or percentages, and measure the quality of something based on a subjective evaluation. Examples are: *the nature of interaction among subgroups of different ethnicities, youth's perception of their role in the community, and the relationship between a teacher and their students.* They are intended to explore and describe judgments, opinions, perceptions and attitudes toward a given situation or subject,⁶ thus aiding an understanding of the participants' experience—what they know, think, like or do—and how that changes over time. Qualitative indicators are particularly suited to complex or nuanced issues or in cases where there is little existing information to provide a basis for quantitative measures.⁷ Like quantitative indicators, qualitative indicators have disaggregation categories.

5. Church and Rogers 2006.

6. CRS and Humentum 2019.

7. Parsons et al. 2013.



Indicators that contain a qualitative dimension, but that methodologically ask for quantitative data collection approaches, will generate quantitative data and are therefore classified as quantitative indicators.

Common confusion around quantitative and qualitative indicators: If an indicator is expressed in quantitative terms that describe qualitative change, is it quantitative or qualitative?

The response to this question varies across the literature and donors. Indicators—such as *percentage of children reporting improved well-being*, *percentage of refugees reporting a high sense of belonging*, and *percentage of women reporting a high sense of agency*—are considered qualitative by some donors.⁸ In this guidance, these are classified as quantitative indicators for several reasons. Although they do contain a *qualitative* dimension that calls for a subjective evaluation by the respondent (e.g., well-being, sense of belonging, sense of agency), they methodologically ask for *quantitative* data collection methods, sampling and tools, and will generate quantitative data. The qualitative “what” within the indicator (e.g., well-being, sense of belonging, sense of agency) is clearly defined prior to data collection—the quality has been “quantified”—and as such is measured through closed-ended questions that generate quantitative data and are analyzed using statistical methods. Indicators of this kind are essential for a quality MEAL system,⁹ especially for measuring changes at the IR and SO levels, since at both levels we expect a certain amount or quantity of change (expressed as a number, percentage, ratio or rate) as well as a specific type of qualitative change, expressed in the “description of what is being measured” part of the quantitative indicator.

For more information on qualitative indicators and frequently asked questions, see [Annex 2](#).

SMART

SMART indicators define the quantity and quality of expected change.

IV. Quantitative Indicators should be SMART

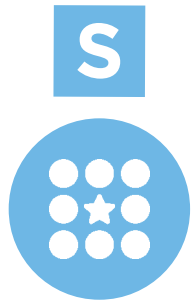
Indicators that are SMART—Specific, Measurable, Achievable, Relevant and Timebound—ensure the gathering of reliable evidence about how much or how well objectives are being or have been achieved.¹⁰ SMART indicators help us clearly define the quantity and quality of expected change, and ensure that it is measurable (i.e., can be reported, counted or observed); that the targets are timebound and achievable within the project’s scope and scale; and that the resultant data will be relevant and useful for decision-making.

8. For example, Global Affairs Canada.

9. People in Need 2021.

10. CRS 2015a.

Figure 1: SMART checklist



Specific

- Is the indicator sufficiently defined so it is clear what is being measured?
- Would two or more project staff members understand it in the same way?
- Are qualitative and/or ambiguous terms in the indicator defined?
- Is the measurement unit specified (including disaggregation categories as appropriate)?
- For percentages, are the numerator and denominator defined?



Measurable

- Can data be collected and analyzed in a timely manner?
- Can the indicator be observed, counted, self-reported or otherwise measured?
- Is the data collection and analysis effort (the expertise, time and staff required to collect and analyze the indicator data) commensurate with the needs of the project?



Achievable (or attainable)

- Is the nature of the change measured through the indicator achievable, given the duration and resources of the project?
- If the indicator collects data that is prone to seasonal changes, can it be collected at the same time of year over the life of the project to demonstrate the change?
- Have the targets been specified? Are the targets attainable given the duration and resources of the project?



Relevant

- Is the indicator the most appropriate measure of progress of the output, IRs and/or SOs?
- Will the data help us gauge whether the output, IRs and/or SOs have been achieved?
- Will the indicator be useful for project management?
- Will the indicator help to test the pathways of change/TOC and inform project learning?



Time-bound

- Does the indicator include information on when the target change is anticipated to take place? (i.e., time qualifier in accordance with the project's activities schedule and TOC)?
- Have targets been set on the relevant timeframes? Is it clearly specified when the target should be achieved?

V. Wording of the Indicator and SMART Dimensions

When looking at the elements of the indicator wording introduced in [Section III](#)—unit of measure, subject of measure, description of what is being measured and disaggregation categories—a natural question comes to mind: Does ensuring the indicator has these elements also ensure it is SMART? The two are interrelated, but not all SMART elements can be included in the indicator wording. However, these elements provide an opportunity to introduce some SMART dimensions into the indicator wording.

Characteristics not contents

SMART does not describe the content of the indicator but its characteristics. Whether or not some SMART dimensions are included in the indicator wording or presented in the supporting design tools, it is critical to run all indicators through the SMART checklist.

Consider the objective statement, *Refugee families properly maintain their temporary shelter*, and its indicator, *Percentage of families occupying temporary shelter who received three or fewer complaints from their neighbors in the past three months*. The indicator could have been written as: *Percentage of families who are considered good neighbors*, which would then require additional work on the definitions (i.e., what does “good neighbors” mean? What would be included in the numerator and denominator of the percentage?). In this case, specificity, or S in SMART, is ensured through the indicator’s wording.

This level of specificity is often easier at lower levels of the Proframe, i.e., outputs. As we move to the upper levels of the RF/Proframe—to the SOs and IRs—a change introduced through the project and captured in the objective statements becomes more complex and often multi-faceted, which in turn makes the indicators measuring those changes more complex. It would often be too cumbersome to incorporate all specificity into the indicator wording. It is recommended that, whenever feasible, specificity be included in the indicator wording, but, if the wording becomes too complex, provide a separate explanation of the definitions and other indicator details in the supporting MEAL documents.

The relevance, attainability and measurability dimensions of SMART are often not obvious from the wording of the indicator and its elements. Note that errors in these dimensions, i.e., a mismatch between objective and indicator (relevance), the feasibility of achieving change measured through the indicator (attainability), the accessibility of the data source and the feasibility of measuring specific change (measurability), may be spotted through a simple review of the indicators and the objective statements they measure. If a time qualifier is relevant, it may be incorporated into the indicator wording, but other aspects of the timeliness dimension are often not obvious and should be included in other MEAL design documents.

The wording of the indicator matters as it may affect how the data for the indicator needs to be collected. Check [Example 1 from the field](#) below.



The way an indicator is worded often informs the questions to be included in the data collection tools.



Example 1 from the field: Wording of indicators and questions in data collection tools

The way an indicator is worded often informs the questions to be included in the data collection tools, especially for indicators that contain qualitative dimensions.

For example, the objective statement: *“Increased capacity of farmers to use new seed selection techniques”* may be measured using one of the following indicators:

1. Percentage of farmers who report a high or very high level of confidence (4 or 5 on a 5-point scale) in using a newly introduced seed selection technique.
2. Percentage of farmers who strongly agree or agree with the statement *“I will be able to use the newly introduced seed selection techniques during the next season”* on a 5-point scale ranging from “strongly agree” to “strongly disagree,” with “neutral” in between.
3. Percentage of farmers with increased capacity in seed selection; “increased capacity” is defined as rating 3 or 4 (on a scale of 1 to 4) to the question: *“How do you assess your capacity to use new seeding techniques now compared to before the training?”* with response options: (a) *it has not changed*; (2) *it has changed a bit, but I still have questions*; (3) *it has increased to some extent*; (4) *it has increased significantly*.

The first indicator uses the question about the level of confidence rated on a 5-point Likert scale. The second uses a slightly different approach: reading a statement and seeking a level of agreement with that statement on a 5-point scale. The third, although at first sight appearing ambiguous and possibly requiring two data collection points (before/after), uses the *definition* to explain how the indicator will be measured and points to the question and ratings to be used in the tool. All three versions are viable options, but each uses slightly different questions in the data collection tools.

Note that scales need to be carefully selected so they are appropriate to the question, data source, management needs and analysis plans. Refer to [Annex 3](#) for tips on using scale-based questions.



Developing indicators is a joint effort between programming and MEAL team members.

VI. Developing Indicators: Who, When and How?

Who?

Developing indicators is a joint effort between programming and MEAL team members. The participation of partner, programming and MEAL staff is essential throughout all steps. Country or regional sector technical advisors should also be involved, especially in the early stages of indicator development, to provide suggestions or feedback on an indicator's feasibility, appropriateness, timing, usefulness, etc. If you have not consulted the sector and MEAL TA during design, be sure to involve them during the review of the MEAL design documents, prior to submission to the donor. While it is useful to get a variety of perspectives when thinking about possible indicators, the refinement and final wording is best done by a small group of key programming and MEAL staff.

Participation in indicator development

Indicator development is the responsibility of programming and MEAL staff. Input by programming staff is essential to ensure that the selected indicators are feasible and useful from a project management perspective, while MEAL staff ensure they are technically sound from a MEAL perspective. Their participation is critical throughout all the steps of indicator development. Meaningful involvement of partners may be challenging given time or capacity constraints, but their input is essential. Partner staff are often charged with collecting the data on the ground and should be considered key users of the information generated by the indicators. Best practice is to accompany partners throughout the process, e.g., work in small groups with MEAL and sector leads co-facilitating indicator design, and organize a joint review process led by programming and MEAL staff prior to proposal submission.

When?

It is recommended that thinking about indicators begins as early as possible in the project design process, ideally at the concept note development stage, without waiting for the TOC or RF to be finalized. While the steps in [Annex 4](#) describe the ideal process of starting with a draft TOC, in some situations you may need to start with just an RF or a Proframe and the implicit TOC that underlies them.

How?

Developing indicators is an iterative process. Some steps presented in [Annex 4](#) may need to be revisited as the team gets more information and their thinking evolves.



Reflect on the indicators using the draft theory of change



Relate the indicators to the types of change/objective statements in the Proframe



Prioritize the indicators with their use in mind



Refine the indicators and run them through the SMART checklist

Are there exceptions to the how-to steps detailed in Annex 4?

Exceptions to the process described below are **immediate emergency projects** that are often short in duration (e.g., 6-12 months) and focus on lifesaving interventions. Most of these projects do not express the target change and assumptions in the form of a TOC, and typically have simple indicators tracking delivery of support, use of delivered items and services, and satisfaction with support. These projects go through steps 2 to 4 in a “telescoped” manner, i.e., the process is much faster, and is based on indicators selected from a menu of standard, tested indicators for emergencies prescribed either by donors or Sphere standards.

How do the donor-mandated indicators fit into the how-to steps?

Donor-mandated indicators are an essential requirement of many proposal submissions.

Donor-mandated indicators are an essential requirement of many proposal submissions. The United States Agency for International Development (USAID) and its Bureau for Humanitarian Assistance (BHA);¹¹ the State Department’s Bureau of Population, Refugees, and Migration (BPRM)¹² and Bureau of Democracy, Human Rights, and Labor (DRL);¹³ and Global Affairs Canada (GAC)¹⁴ are among the many donors that require inclusion of their own indicators, some that are mandated and others that are strongly recommended.¹⁵ This is because each donor agency has its own constituencies to whom it needs to report, and therefore requires that implementing agencies use indicators to capture progress in each of the sectors and technical areas it funds. This set of indicators, called **standard indicators**, is to be given priority over alternatives wherever a particular standard indicator would be applicable.¹⁶

11. See [Emergency Application Guidelines, Annex B: Indicator List](#) (USAID BHA 2020a) and [Indicator handbook for emergency activities](#) (USAID BHA 2020b).

12. See [General NGO guidelines](#), Appendix C: Programmatic Sectors, Modalities, and Standardized Indicators (BPRM 2021).

13. See [Proposal Submission Instructions for Applicants](#) (DRL 2020).

14. See [Feminist International Assistance Policy Indicators](#) (GAC 2016b).

15. For examples of other sector-specific standards and donor-mandated indicators see: Food for Peace Indicators Handbook, [Part I](#) (USAID 2020a), BHA Indicators Handbook [Part II](#) (USAID 2021a) and Food for Peace Indicators Handbook [Part III](#) (USAID 2020b), and [Global Reference List of 100 Core Health Indicators](#) (WHO 2015).

16. USAID [Standard Indicators](#).



Industry standard indicators are often accompanied by detailed guidance.

Donor-mandated and other indicator definitions

- ◆ **Standard indicators** are also called donor-mandated indicators. They are listed or referenced in the donor's request for proposal. They are often sector-specific and must be included in the project proposal.
- ◆ **Industry standard indicators** are indicators that specific sectors identify as best measures for interventions in those specific sectors, often developed collaboratively by practitioners. Examples of these are: Sphere Standards,¹⁷ detailing a set of principles and minimum humanitarian standards for water supply, sanitation, hygiene, shelter, etc.; or a set of food security indicators, including Food Consumption Score, Household Hunger Scale, Coping Strategy Index, Household Dietary Diversity Score, Women Dietary Diversity Score, etc.¹⁸ These are often accompanied by detailed guidance explaining how the data collection tool should be developed and/or adjusted to the local context, and how the data should be collected and analyzed. As they are based on industry knowledge and have often been tested in many contexts, donors frequently include them in their list of standard indicators.
- ◆ **Globally accepted indicators (GAIN)¹⁹ or indicator banks for specific sectors²⁰** CRS uses the term GAIN to denote industry standard indicators in areas in which there is no guidance and that CRS is gradually building experience in. Similarly, many sectors are developing indicator banks listing sample indicators and indicator tables developed based on reviews of the relevant indicator literature and CRS projects. Note that the indicators presented in the indicator bank are often illustrative of the types of indicators that could be used to measure change, but have not necessarily been tested or validated.
- ◆ **Custom indicators** is a term used by some donors to refer to indicators that an organization can design and/or select itself. These can be selected from the donor's standard indicator list or industry standard indicators. These are indicators that are customized to the project.

17. [Sphere Humanitarian Standards](#).

18. Maxwell et al. 2013.

19. In 2008, CRS began the GAIN initiative to strengthen indicator practice across sectors. As a result, GAIN indicators for the [peacebuilding sector](#) have been developed.

20. See [Social Cohesion Indicators Bank](#) (CRS 2019).



VII. Tools for Development of Indicators

There are a number of tools that help develop indicators, each useful at different stages of their development.

There are a number of tools that help develop indicators, each useful at different stages of their development. Every detail in the indicator matters and it is always preferable to be as clear as possible at the time of project design to avoid challenges in the future. Documenting the indicator definitions and rationale behind MEAL design decisions ensures that critical information and assumptions are not lost during handover from the project design team to the implementation team. Also, most donors ask for a detailed presentation of the indicators, including clear definitions of what is being measured and how; how information will be analyzed and used; and the roles and responsibilities throughout the process. The tools most frequently used to help think through these questions or present the results of the design process are shown in [Figure 2](#) below, together with details of the key information they include, when they are most useful, and tips and best practices.

Figure 2: Overview of MEAL design tools

Tool	Format and information contained	When to use it	Tips and best practices
Proframe or Logframe ²¹	A 5 x 4 matrix including objectives (goal, SOs, IRs, outputs and activities), performance indicators, measurement methods/data sources and critical assumptions. It provides a snapshot of the project's key design and MEAL elements.	Most useful at the initial design stage, after a project's RF has been drafted, as the design team starts planning for the project's MEAL elements. Most donors do not require it.	The only tool that contains assumptions. The ProFrame is best used as an early project design <u>tool</u> that guides the project design team as they think about the objectives hierarchy, assumptions, if-then logic, and associated indicators and monitoring plans. It forms the basis for subsequent development of the performance monitoring plan (PMP). ²² There is no point in developing a Proframe after the project has been submitted.
Performance monitoring plan²³ (PMP) or MEAL plan	A table that includes: <ul style="list-style-type: none"> ▪ Indicator, including unit of measure, disaggregation categories, definition of ambiguous terms, and formula for calculation (if applicable). ▪ Data collection details, including data source, method, tool, sampling, who is responsible, and how often data is being collected. ▪ Data analysis, including type and/or approach to the analysis, frequency, and responsibility. ▪ Data use describing how specific information about the indicator will be used for evidence-based decision-making within the project. 	Often part of the proposal submission required by most donors. Useful for fleshing out indicator details.	Start working on the PMP only when the indicators are close to final. If a donor requires both a PMP and a PIRS, the PIRS may be helpful at an earlier stage to guide the teams as they think through the details of each indicator. If you need to submit both a PMP and a PIRS, keep the PMP focused on the major elements of the indicator, with most details documented in the PIRS. This will help keep the PMP short, and easy to read and use. Like the ProFrame, the PMP helps teams understand how the complement of indicators matches the TOC / project information needs, while the PIRS provides in-depth details on each individual indicator – see below.
Performance Indicator Reference Sheet (PIRS)	The PIRS requires similar information to the PMP, but in greater detail, and includes some additional information, e.g., limitations of the indicator or data collection or both, plans for data quality assessment, etc.	If required by the donor at the proposal stage, it is developed in parallel with the PMP and submitted with the technical narrative. If not, it should be drafted either during the project design stage to capture early thinking, or during start-up, after the project is awarded. In both cases, the PIRS is validated and refined during the SMILER+ session.	While there is some overlap between the PMP and PIRS, the PIRS format focuses on individual indicators, allowing documentation of all their details. If required by the donor, a PIRS is developed for all indicators. If not required by the donor, it is strongly recommended that a PIRS is developed for all SO- and IR-level indicators, as well as for any complex indicators at the output level for internal reference. The PIRS is useful for internal documentation and institutional memory even if developed after the project is approved. For an example of a PIRS template, instructions, and tips, refer to Annex 6 .

21. For more information on the Proframe, refer to *ProPack I* (CRS 2015a).

22. Adapted from *ProPack I* (CRS 2015a) and *A guide to the MEAL DPro* (CRS and Humentum 2019).

23. A performance monitoring plan (PMP) is also referred to as an M&E plan, MEAL plan or performance monitoring framework.

Many donors allow, and in some cases even encourage, treating the PMP as a working document. However, projects should not change indicators without strong justification for doing so, since this may hinder the team's ability to compare results over time and to thereby measure a change over the life of the project.²⁴

Always check the donor requirements and templates for the Proframe/Logframe, PMP and PIRS. Each donor uses a different format and asks for different levels of detail. Most donors provide instructions on what they want to see in each of these tools.

VIII. Common Indicator Challenges, Good Practices and Tips

Poor work during any of the indicator development steps can have severe implications for the development of the MEAL system.

While the process for developing indicators presented in [Annex 4](#) works for all types of projects that have a TOC and/or RF/Proframe, poor work during any of the steps can have severe implications for the development of the MEAL system and its further use for project management and learning. In practice, challenges with indicators can be due to a number of factors related to the SMART dimensions. Note that, given the strong interconnectedness of the SMART dimensions, resolving challenges in one dimension may have effects on another, e.g., the more specific the indicator is (S dimension), the easier it may be to measure (M) and the more relevant (R) it becomes for project management. Meanwhile, adding a time dimension (T) may affect the indicator's attainability (A).

24. For more information on the PMP template and how to fill it out, refer to [A guide to the MEAL DPro](#) (CRS and Humentum 2019), page 42.

S M A R T



SPECIFICITY: TYPICAL CHALLENGES AND IMPLICATIONS

Challenge: Indicators lack specificity

Indicators may be phrased as a mere restatement of the objectives without clearly defining the change we expect to see. These might include ambiguous or complex terms such as “sense of belonging,” “well-being” or “resilience,” or qualifiers such as “adequate,” “safe,” “quality” or “Sphere compliant.” This challenge is especially prevalent at the higher levels of the RF and TOC that capture more complex changes resulting from the intervention. The definitions become either too theoretical or too complex to result in useful information for project management.

Definitions of complex or composite, multidimensional indicators often go into unnecessary detail, separately defining each dimension of the indicator but failing to define the overall concept that is being measured and how all terms or dimensions in the indicator work together.

When definitions are poorly explained, there may be ambiguity about what specific data needs to be collected.

Additionally, indicator definitions often include ambiguous terms (e.g., participation is defined as “engagement”) or the same term is defined differently in indicators across different levels. When definitions are poorly explained, there may be ambiguity about what specific data needs to be collected, resulting in many irrelevant or “nice-to-know” questions in the data collection tools and overwhelming data analysis and interpretation.

Another typical challenge relates to omitting the formula for calculation in the indicator definition, especially for indicators expressed as a percentage. If the numerator and denominator are not clearly specified, this may lead to errors in specifying targets or difficulties when trying to determine the denominator and calculate the indicator (e.g., you may need a population-based survey to collect the data required for the denominator).

Finally, all possible disaggregation categories are often included rather than only those most useful from a program management perspective.



Example 2 from the field: Complex indicators

A peacebuilding project uses a 3B approach—binding, bonding and bridging—to bring about the following strategic objective, to be measured through the accompanying indicator:

Strategic objective	Indicator
Diverse tribal, religious and ethnic groups in the targeted area have increased mutual understanding, tolerance and trust.	Percentage of participants who have increased mutual understanding, tolerance and trust.

The indicator is simply a restatement of the objective. This may not necessarily be wrong if the definition (1) clearly defines each ambiguous term—“mutual understanding,” “tolerance” and “trust;” (2) determines how the three work together; and (3) includes a formula to calculate this percentage.

This may become very complex. How much of an increase in each of the terms is required to qualify as “increased” within the aggregate indicator? Should all the dimensions be given equal weight in the overall score or are some more critical than others in the project context?

In practice, each of the three terms have a multi-dimensional definition, which will require several questions in the data collection tool to get the complement of information for each dimension, which then needs to be aggregated to arrive at the overall indicator inclusive of all three dimensions. The formula for calculation can become very complex as it requires a way to score responses across each dimension, and then define the minimum score required for a respondent to be counted in the numerator of the percentage calculation (e.g., What qualifies as “increased” understanding? Does any increase over baseline count? Should it occur in all three dimensions, or would a positive change in one or two suffice? What if a participant’s score is higher in one dimension but lower in another? etc.).

S M A R T



SPECIFICITY: GOOD PRACTICES AND TIPS

Good practice: Define all ambiguous terms in the indicator but avoid unnecessary detail.

Definitions should be locally meaningful and simple.

Definitions help clarify the indicator—and, by proxy, the objective—thus showing how it is relevant to the project. The basis for the definitions usually comes from Steps 1 and 2 of the indicator-development process described in [Annex 4](#). (Recall the tip to use detailed explanations of the change you want to see at each level of the TOC and objectives hierarchy, reflecting the assessment findings as prioritized during the problem analysis). The definitions should be developed by programming staff, with input from technical/sectoral experts and, if appropriate and feasible, community members (see [community-defined indicators](#) below). The definitions should be locally meaningful and simple, focusing on one dimension within the indicator. The same term should not have different definitions across different indicators of the same project. For example, if an indicator at the output level measures number of constructed “safe” places for youth, and an indicator at the IR level measures percentage of youth satisfied with “safe” places, “safe” needs to have the same definition in each indicator. Avoid defining every word in the indicator.



Example 3 from the field: Improving indicator definitions

The objective statement *Boys and girls in areas of displacement or return have access to a safe physical school environment that meets Inter-agency Network for Education in Emergencies (INEE) standards²⁵* is measured by the following indicator: *Number of assessed formal and informal learning environments that are considered safe for boys and girls of different ages*. During the initial development of indicator definitions, the team came up with the following:

- ◆ **Assessed:** Rated using a scale from “meets requirements” to “does not meet requirements” that is then transformed into a score.
- ◆ **Formal and informal learning environments:** CRS- and Caritas-supported government schools used during the academic year and for summer classes.
- ◆ **Considered safe:** Obtains a score of 75% or above on the Quality Learning Assessment (QLA).
- ◆ **Safe:** Safety and security meet INEE standards and comply with Department of Education/municipal regulations.
- ◆ **Different ages:** Students in primary school (aged 5-13) and secondary school (aged 14-18)

25. INEE defines standards for the minimum level of educational quality and access in emergencies through to recovery.

SMART

In this case, the team approached the definition in a word-by-word manner, defining almost every word mentioned in the indicator. The definitions of some terms partially overlap, providing unnecessary detail, e.g., the term “safe” has been defined twice, while the term “QLA” has not been defined. The ages of students don’t need to be defined, as the age focus is likely explained in the proposal or can be added as a footnote since it applies across all indicators.

This is how the team improved their definitions:

- ◆ **Formal and informal learning environments:** CRS- or Caritas-supported government schools used during the academic year and for summer class.
- ◆ **Considered safe:** Obtains a score of 75% or above on the QLA assessing physical school buildings and furniture standards.
- ◆ **QLA:** Tool that assesses school safety and security based on INEE standards and Department of Education /municipal regulations, adjusted to requirements for primary and secondary schools.

For composite indicators, be sure to first break down their component parts and then bring them all together in the definition.

For composite indicators, be sure to first break down their component parts and then bring them all together in the definition. If the indicator includes several complex terms, as in Example 2 above, break down and define each component separately, even though you may have to report to the donor on the aggregate result. Knowing how the project progresses against each dimension will be essential for adaptive management. In general, composite indicators require a more nuanced interpretation of results, e.g., highlighting shortcomings in a specific dimension and thus enabling the project to devise an action plan to address these. Even though the donor-approved indicator may be an aggregate, composite indicator of all dimensions, the donors will welcome this more detailed analysis and interpretation and resultant recommendations.

Avoid composite indicators with too many dimensions

To prevent the indicator definition and then the analysis becoming an overwhelming mathematical exercise, avoid composite indicators with too many dimensions. When deciding on and defining indicator dimensions, focus on the most problematic or the most important issues identified in the assessment. If you are uncertain, build language into the proposal to finalize this decision after the baseline data has been collected. Note that this does not mean the team cannot or will not investigate additional dimensions, but they will focus on the most severe/problematic ones for reporting as a proxy of overall change.

You may also negotiate with the donor to allow reporting on each dimension separately to avoid weighting and to simplify the calculation, and enable a more nuanced understanding of the results across all dimensions, which might otherwise have been hidden in the aggregate’s average. In practice, the data on all dimensions is already being collected, and this option significantly reduces the calculation effort and the chance of a calculation error. At the same time, it allows an investigation of potential correlations during the analysis, without introducing complex formulas.

SMART

If, however, the donor requests reporting on the aggregate indicator, it is essential to define how all dimensions of the entire concept being measured work together to generate the indicator, specifying the possible weighting of some dimensions over others, or what or how much change in each dimension will qualify for counting a respondent in the numerator.

If you are considering weighting each dimension of a composite indicator, it is essential to consult the sector technical advisor.

Weighting each dimension

If you are considering **weighting each dimension of a composite indicator**, it is essential to consult the sector technical advisor. For example, in agronomy interventions, there may be several behaviors involved, each of which affects productivity in different ways and to a differing degree. The indicator is often *Percentage of farmers who adopt three of the five proposed practices*.²⁶ One or two practices might account for most of the productivity increase measured at the SO level. However, counting three practices that had been adopted—but that had accounted for little productivity—may be mistaken for success at the IR level. To avoid this, it is essential that the sector TA is consulted to establish which behaviors are key to the SO-level change you expect. It is also critical to consult assessment results to confirm that these behaviors were not already practiced before the project began.

While it is important to be as specific as possible in the wording of the indicators and their definitions in internal design documents, you may want to share less detailed information or present these proposed definitions as illustrative in the donor submission. This will ensure CRS is not contractually bound to follow a very precise approach until it has been field tested to confirm that it is appropriate, which typically happens during baseline data collection.

Good practice: If you are uncertain how to define an indicator, consider asking communities.

Communities know best what kind of change they want to see as result of our work. Consulting them is an important accountability mechanism. These consultations are often needed to measure high-level objectives, i.e. SOs. Community input may come in two forms:

- **Community-defined indicators**²⁷ – when an indicator is not developed. Community-defined indicators entail light formative research with in-depth qualitative data collection before or at the beginning of the project. The questions asked of the communities are open-ended and aimed at understanding how target participants most directly affected by the issue would define project success. When communities define project success in their own words, this becomes a community-defined indicator. It is often “quantified” or turned into a quantitative indicator that is measured through series of questions using a Likert or similar scale, with a clear calculation procedure.

26. Adoption of even simple behaviors can take many years. During a project of two years in duration, farmers are trialing and adopting. While our indicators often focus on measuring only adoption, in projects that include social behavior change, it is important to capture both via data collection tools.

27. For more information, see [Community-Defined Indicators Guidance Sheet](#) (CRS 2015b) 

S M A R T

- **Community input on defining ambiguous terms in your indicator** – when an indicator is developed, but does not define ambiguous terms that are complex and need to be contextualized. Examples of such indicators are: *Percentage of project participants reporting improved resilience* or *Percentage of women reporting a high level of agency*). In this case, a small sample of target participants is consulted to help define the ambiguous terms. This effort does not start with a blank paper and open-ended questions to community members; instead, teams generally use this process to probe assessment findings, refine their initial understanding, specify expected end-of-project change, and define the indicator with as much nuance and specificity as possible. A good practice is to first review secondary data to understand the industry definition of the term and then, in consultation with communities, contextualize that definition.

Community-defined indicators are recommended only for higher-level indicators.

When formulating community-defined indicators or soliciting community input on indicator definitions, it is recommended that men and women are consulted separately, to get both perspectives and incorporate the potential differences in men's and women's definitions. The most frequently used data collection method is focus group discussions (FGDs), to ensure rapid collection of multiple perspectives on the topic under investigation and capitalize on rich insights coming from interaction among group participants.²⁸ The data collection does not need to be heavy; a few FGDs with each subgroup should suffice.

Field testing essential

Field testing is the best way to validate indicator definitions and check whether the data generated through the tool will enable the correct calculation of the indicator.

When the community-defined definition is not specific enough, **the baseline can also help refine or validate the indicator definition**. The baseline should not only focus on quantitative data and indicators; it is also an opportunity to identify and clarify any issues with indicator definitions, which could lead to a more nuanced understanding of ambiguous terms and indicator refinement. Community-defined indicators or seeking community input into indicator definition are recommended only for higher-level indicators, not for indicators measuring IRs or outputs, as the expected changes and definitions should be clearer at these levels. For more information on tips and challenges with communities providing input into definitions, refer to the case study in [Annex 7](#).

Good practice: When an indicator is expressed as a percentage, include a formula for calculation.

The formula needs to include the unit of measure, numerator and denominator, taking into consideration the proposed sampling approach. The wording of the indicator may have implications for the formula as well as for the frequency of data collection and its overall usefulness for project management.

28. For more information on how to design and conduct FGDs, refer to [Practical guide to focus group discussions](#) (Dzino-Silajdzic 2020).

Percentage calculation conundrum

The phrasing of the indicator influences the formula for calculation, as well as data collection frequency, and the sampling and usefulness of the indicator for project management. Let us examine this through the examples below, typical of projects that include outputs about increased participant knowledge as a result of project-delivered trainings. Indicators for such outputs typically aim to measure a change in the knowledge of training participants. This is also an example of the interconnectedness of SMART dimensions as the actual increase measured in examples 1, 2 and 4 greatly depends on the baseline level of knowledge, thus affecting the "attainability" dimension of SMART.

Indicator: Percentage of participants with increased knowledge of topic X.

Formula:

$$\% = \frac{\text{\# of randomly sampled training participants that meet the condition: [if (posttest score - pretest score) > zero]}}{\text{total \# of trained participants in a sample}}$$

Implications: "Increased knowledge" requires two data points for the sampled participants (pre- and post-training score) and a comparison of the results between the **same participants** in the pre- and post-training tests. The indicator is not very specific and may not be very useful for project management: How much of an increase in knowledge would count? How does this compare with our assumptions about the minimum knowledge change that may be required for our pathway of change or TOC?

Indicator: Percentage of participants with at least a 10% increase in knowledge of topic X.

Formula:

$$\% = \frac{\text{\# of randomly sampled training participants that meet the condition: [if (\frac{\text{posttest score} - \text{pretest score}}{\text{pretest score}}) \geq 10\%]}}{\text{total \# of trained participants in a sample}}$$

Implications: A double percentage calculation is required, as well as two data points for the **same sampled participants**. This is unlikely to be calculated correctly and the question is whether it is worth the effort. Try to avoid phrasing an indicator in this way.

Indicator: Percentage of training participants who scored at least 75 points (out of 100) on a knowledge test on topic X.

Formula:

$$\% = \frac{\text{\# of randomly sampled training participants that meet the condition: [if test score \geq 75 points]}}{\text{total \# of trained participants in a sample}}$$

Implications: It requires only one data point/one data collection effort (post-training test). We may still administer a pre-test and compare the results for our internal use, but we are not obliged to do so, **nor do we need to sample the same respondents**. In most of our projects, a precondition at the output level is that participants acquire enough knowledge to perform a desired task or behavior (regardless of how much knowledge change this represents). This indicator is usually sufficient and is more useful for project management purposes. This option is not only the easiest to manage but is also likely the most powerful as it enables further probing into questions that may have been answered correctly or incorrectly in the post-test, and then addressing knowledge gaps through adaptive management actions.

Indicator: Percentage increase in knowledge on topic X.

Formula:

Calculate total knowledge score in pre- and post-training test and then average score across all sampled training participants.

$$\% = \frac{\text{posttest average score} - \text{pretest average score}}{\text{pretest average score}}$$

Implications: We will learn the aggregate knowledge level before and after the training, and the percentage change between the two, but we will not know the number or percentage of individuals affected or how many may have reached the knowledge level required to succeed. This indicator **does not require the same participants to be sampled** and is easy to calculate, but may not be very useful for project management purposes.

Similar thinking can be applied to higher-level indicators that measure resilience, tolerance, etc. None of the options are wrong from a mathematical perspective; however, each has implications for data collection, sampling and the usefulness of the information generated by the indicator. Note that sampling requirements may have a significant impact on resources and on how the data collection is organized. Always be sure to use phrasing and percentage calculations that generate the most useful data for project management. Remember that we should be aiming for the simplest approach to meet reporting requirements, but if information needs for project management require additional data collection, i.e., pre-test in Example 3, nothing prevents us from doing so.

Should we always avoid the words “increase,” “decrease” and “improve” in indicator formulation?

No! Having indicators that measure improvement or increase may be appropriate in the following cases:

- When we need to use a **donor-mandated indicator** and are not permitted to change its wording.
- When we need to measure a **change at the individual level**, for example in access, understanding or behavior. This is often the case in peacebuilding, social cohesion or gender-transformative projects, aimed at changing norms and behaviors.
- When the **amount of increase or improvement does not matter for project logic**, when any increase counts. This often happens in livelihoods projects when we measure income or production.
- When there is **no acceptable minimum standard**. If a minimum standard exists, it is much better to phrase the indicator as meeting this standard. For example, an indicator measuring the *percentage of households with improved access to a water source* would be more SMART and easier to calculate if phrased *percentage of households reporting daily access to a water source in the previous seven days at the time of data collection*.

Note that even when you cannot avoid using words that denote the direction of change in indicator wording, there is still a way around it to avoid challenges related to sampling the same respondents. In this case, it is recommended that the word “increased”, “improved”, etc. be used in the definition, to illustrate a status, level or situation at the specific time of data collection. For example, in the indicator *Percentage of targeted youth with improved tolerance*, “improved tolerance” may be defined as: rating “strongly agree” and “agree” with statements 1, 2 and 3. This means that there could be a certain percentage of youth that would be qualified as having “improved tolerance” even at baseline. In this approach, we do not need to sample the same participants as we are not focusing on tracking individual participants’ change. We are rather tracking a change at the cohort level – if the percentage of youth with desired state of tolerance increased or decreased. See *Example 4 from the field* below to learn more about this approach.

S M A R T

Good practice: Make the donor's indicators work for you.

Define any broad or ambiguous terms in the indicator so that they reflect what your intervention focuses on.

While some donors have specific indicator definitions, and requirements related to data collection methods, sampling and calculations, whenever possible, adjust or further define donor indicators so that they **meet your and other project stakeholders' information needs**.²⁹ The best place to do so is in the indicator's definition, although some adjustments may be made to the methods of data collection, analysis, etc. Define any broad or ambiguous terms in the indicator so that they reflect what your intervention focuses on, and include key changes that need to be measured at each level of the objectives hierarchy. See [Annex 5](#) detailing steps in adjusting donor-mandated indicators.

**Example 4 from the field: Adjusting donor indicators**

The Bureau of Population, Refugees, and Migration requires all projects to include this standard outcome indicator: *Percentage of beneficiaries who report an improved sense of safety and well-being at the end of the program, disaggregated by age and gender*.³⁰ BPRM guidelines suggest the following definitions:

- ◆ “Sense of safety” is defined as “A reasonable age- and circumstance-appropriate level of comfort and satisfaction in security levels as well as a lack of fear in surroundings (i.e., in and around places or individuals).”
- ◆ “Sense of wellbeing” is defined as “A general sense of either comfort, trust, health, mental/psychosocial stability, functionality, and/or freedom from abuse, neglect, exploitation, or violence.” The definition of well-being should be context-specific and locally validated.

A CRS team working on a livelihoods small enterprise project adjusted the definition to:

- ◆ Sense of safety: Respondents' perception of how safe it is to operate their business – (illustrative statement: *“I feel safe to operate my business in [city]”*).
- ◆ Sense of well-being, defined through two sub-components deemed most critical for this project:
 - Optimism – (illustrative statement: *“I feel optimistic about the future”*)
 - Resilience – (illustrative statements: *“I am able to deal with problems well”* and *“I feel I have a support network to help me through the challenges”*).

The indicator was measured using a five-point Likert scale with ratings from “strongly disagree” to “strongly agree,” and a neutral midpoint. Respondents were counted in the numerator—as having an “improved sense of safety and well-being”—if they expressed agreement (“agree” or “strongly agree”) with all four statements. This is an example of incorporating the direction of change—“improved”—into the indicator definition to avoid sampling the same respondents throughout all data collection efforts and calculating whether each sampled individual had achieved change. From a TOC perspective this worked well, as the team learned during the assessment that safety and well-being components chosen as focus in the indicator definition were generally low and most critical to focus on. By tracking the cohort-level change, they can safely assume that individual change is also taking place.

29. Consult your regional technical advisor for business development to understand how much flexibility the donor may allow you when using its mandated indicators.

30. BPRM 2021, Appendix C.

SMART

This indicator still requires a baseline–endline comparison at minimum. In the case of outcome-level indicators such as this, especially for projects that last for two to three years, we may see only minimal improvements between baseline and endline. Therefore, good practice is to plan to calculate several measures in the analysis plan, in case the specific measure identified at MEAL system design does not show a significant change. Apart from average well-being ratings and binary comparisons (i.e. combining all “positive” and “negative” responses together), you will also want to calculate and compare the percentage of respondents in each category of response option, to evidence how the percentage of responses shifted between different response categories. For instance, the change in the percentage of respondents who agreed or strongly agreed to all four statements might be small, but there could be a significant shift in responses from “agree” to “strongly agree” or there may be significant increases in just two of the four dimensions of well-being. Both of these would in fact be reasonable signs of improved well-being, even if this was not the specific definition proposed by the team at the MEAL system design stage.

Good practice: Use disaggregation categories to meet project management and learning needs.

Collecting and recording data by referring to specific demographic or vulnerability characteristics can increase the usefulness of an indicator.

Collecting and recording data by referring to specific demographic or vulnerability characteristics—disaggregating data by disaggregation categories—can increase the usefulness of an indicator by revealing levels of achievement and trends for relevant subgroups that are important to the success of an activity or project.³¹ Disaggregating data during analysis invites comparison across subgroups, which can enhance an understanding of programming successes and challenges. For example, disaggregating data on boys’ and girls’ school enrollment by location may reveal locations in which the project is more or less successful, as well as potential gaps related to girls’ access to education. Such comparisons and accompanying discussions and interpretations are most useful for project management.

Planning for disaggregation

Discussing disaggregation at the time of indicator development and before data collection is critical, as it may be extremely challenging or impossible to reconstruct the data later. The choice of disaggregation categories is often informed by the assessment. During the assessment we learn about subgroups of the population that are potentially more vulnerable, or discover subgroups that we assume would experience different project results, e.g., single-headed families versus families with both spouses. This information points to disaggregation requirements once the project is implemented. While we can disaggregate based on every characteristic of a measurement unit, the decision on disaggregation categories is based on **the assumption that results can be different for specific subgroups** (e.g., between locations with different environmental factors, men and women, people with different breeds of ewes) **or the need to document that they are not**. Ensure that all data collection tools and reporting templates reflect your plans for data disaggregation.

31. USAID 2021b.

SMART

Don't forget that **disaggregation categories need to be associated to the unit of measure** you use in the indicator. If the unit of measure is a household, you cannot disaggregate by gender, only by the gender of the respondent (on behalf of the household).

Too granular disaggregation may jeopardize project participants' privacy.

Remember that **too granular disaggregation may jeopardize project participants' privacy**. Be cautious when collecting personally identifiable information on individuals, such as their ethnicity, religious affiliation, disability, marital status or occupation. These details, if revealed in combination (e.g., marital status, age and occupation), could result in negative consequences in a given context if they enable the individual to be identified. To protect project participants' privacy,³² consider using disaggregation “buckets,”—e.g., age ranges, or district and town rather than specific addresses—and ensure that disaggregation categories remain large enough to allow for meaningful interpretation.

Best practice—as required by some donors—is **setting targets per disaggregation category**. Remember, the more disaggregation categories there are, the more demanding will be the work to set the targets. Focus only on disaggregation categories that are useful and necessary for management decision-making.

Use of disaggregated data in project management

In addition to **comparison across subgroups**, disaggregated data should also be **compared to the baseline and previous performance periods** to check for changes over time in each specified subgroup, and to compare the extent of achievement **against targets** for each specified subgroup. Such analysis can reveal performance gaps for specific subgroups that could be undetected in overall calculations. For example, through a review of disaggregated data, the team may notice that one subgroup is doing worse than expected and when compared with other subgroups. This could prompt additional qualitative data collection to investigate why and whether adjustments to the activities are required to ensure no vulnerable subgroup is left behind.

Don't forget that comparisons between disaggregation categories require **different numerators and denominators**. For example, consider the indicator *percentage of project participants who reported that cash assistance helped them meet their household needs*, disaggregated by gender. If the data was collected from 100 respondents, of whom 30 were male and 70 female, to calculate disaggregated values, we need to use different denominators for each subgroup. E.g., for men, we would be calculating the percentage using the number of men who said cash helped them meet their needs divided by the total number of men, in this case 30.

32. Cobo et al. 2019.

S M A R T

Further disaggregation options may arise during project implementation and may be critical to testing the TOC and capturing learning.

Use of disaggregated data to enforce project learning and test the TOC

Disaggregation categories typically relate to demographic characteristics such as gender, education level, age and vulnerability. However, further disaggregation options may arise during project implementation and may be critical to testing the TOC and capturing learning. Our projects often rest on the premise that if targeted groups follow certain practices, their situation will improve. If, during project implementation, you realize that some project participants have adopted the promoted practices (the so-called “doers”) while some have not (the “non-doers”), the analysis and comparison of the results between these two subgroups, called a **doer/non-doer analysis**, can offer key insights into whether our TOC is being upheld or not.

For example, consider comparing levels of adoption of selected practices between farmers who attended all the training sessions versus those who attended only a few. If our pathway of change from output to IR is correct, the level of adoption should be higher among those who attended all of the training sessions. Such an approach to analysis can be applied to the higher-level TOC and to compare achievements at the SO level too, i.e., increases in livestock productivity between farmers who adopted the recommended practices (“doers”) and those who did not (“non-doers”). This type of analysis and disaggregation that examines related levels of the project’s RF is **a powerful way to test a project’s TOC**. It provides evidence of causality between project strategies and results achieved or, in other words, determines attribution, without requiring a population-based survey or experimental or quasi-experimental design with a control group.³³ A doer/non-doer analysis does not pose the ethical concerns that control groups may, as adoption of practices is a project participant choice rather than a programming-controlled factor.

The presence of doers and non-doers needs to be confirmed through ongoing rigorous or light monitoring.³⁴ For example, adoption of the practices may be checked through observation or monitoring of lower-level indicators. Whether you will have doers and non-doers may also be known from previous experience, e.g., lessons learned from previous projects showed that certain groups or locations traditionally did not quickly adopt new practices. Note that if there are no non-doers, you will need to consider other options.

33. A population with the same characteristics as the project participants, but that *has not* benefited from the project. In traditional experimental design, outcomes for the population targeted by the project are compared with a control group to determine the difference made by the project (Hind 2010).

34. For more information, refer to [Practical guide: Monitoring for problem-solving, adaptive management, reporting and learning](#) (Morel et al. 2019).

S M A R T

Doer/non-doer disaggregation and analysis is also a **creative solution to challenges related to baseline/endline comparisons** when such comparisons bring little or no expected changes in indicators. This happens in situations when:

- ◆ Project results are highly influenced by the broader context, such as climatic influence on agriculture and food security outcomes.
- ◆ Baseline results are not reliable, for example, they may be too high due to self-reporting bias.
- ◆ When the project did not conduct a baseline in the first place.

When doers and non-doers can easily be identified, a doer/non-doer analysis is a reliable approach to demonstrate the project outcomes and TOC.

S M A R T



MEASURABILITY: TYPICAL CHALLENGES AND IMPLICATIONS

Challenge: Indicators are not measurable

The feasibility of the proposed data collection method in terms of resource requirements and accessibility of the data source is the most common measurability challenge. The budget available for data collection and the organizational capacity to access and engage with a specified data source play a significant role. For example, indicators that measure percentage increase or improvement may be very time- and resource-intensive, as they require sampling of the same respondents. In some cases, this may be impossible due to population movement or participants dropping out of the project. Also, we may not be able to measure, for example, *percentage of students demonstrating improved learning outcomes*, as the official school records are confidential documents, and we may not get permission from the authorities to access them. At the same time, we may not have the time and resources to perform the tests ourselves, even on a sample of students. Another common challenge is that indicators may contain unrealistic qualifiers and therefore are impossible to measure, e.g., *percentage of households reporting year-round food security*. While this may be appropriate wording for an objective statement (e.g., *Targeted households enjoy year-round food security*), it is only realistic to expect respondents to recall behaviors—such as food security practices—over a week or, at most, a month. Recording food security over a whole year would require a huge level of effort to collect data that is not likely to be particularly useful. Indicators may also measure something unrealistic to achieve or measure within the project duration. This is often the case with projects tied to seasons, e.g., agriculture or school year and semesters in education.

Indicators may contain unrealistic qualifiers and therefore are impossible to measure.

In some contexts, otherwise typical indicators simply cannot be measured with appropriate rigor or quality. For example, when working with illiterate populations, knowledge checks through pre- and post-training tests are less feasible or the data source may be reluctant to take it. Sometimes they simply may not be viewed as appropriate, e.g., in one project, government officials were unwilling to take a knowledge test. A project may be too short in duration to accommodate midterm or endline surveys that could include knowledge check questions. In these cases, there is no easy way to measure participants' knowledge or key message retention in a rigorous manner, yet knowledge gain is often a primary pre-condition for behavior change at the IR level. Even when feasible, the level of effort required may not be justified.

SMART

Data collection for specific indicators may pose ethical or protection concerns, e.g., checking the percentage of women reporting abuse by their family members may put women at risk of further harm, or checking ethnic tolerance in a post-conflict environment may be too sensitive or risky. This issue is often encountered in projects that are trying to address social or cultural taboos, and something that may work in one context may be unmanageable in another. The implication of these challenges can be severe, as we will simply not be able to capture whether the intended change has taken place and ultimately will not be able to report on that indicator.



Example 5 from the field: Are you measuring the unmeasurable?

A 12-month project implemented in five communities aimed to ensure that children had improved access to pre-school education by ensuring a safe and appropriate learning environment (IR1) and the use of inclusive teaching methodologies by teachers (IR2). The SO and proposed indicators at this level were:

Strategic objective	Proposed indicators
Improved equitable access to quality kindergarten education that is age-appropriate, inclusive and gender-sensitive.	Percentage of children in targeted communities enrolled in kindergarten Percentage of children reporting satisfaction with quality education

While both indicators at first sight seem to fit well with the proposed intervention, both will likely be very challenging for the project team.

For the first indicator, the team will require the entire number of kindergarten-age children in each community the project operates in as the denominator for the percentage calculation. If there is no reliable and up-to-date secondary data of this kind, reporting on this indicator would require an enormous level of effort and resources.

The second indicator calls for children as data sources. Children in kindergarten are typically younger than 7 years of age. Practice has shown that children that young cannot participate in an introspective assessment of their own experience, given their cognitive stage of development, thus the data would not be deemed reliable.³⁵ While they could and should be consulted through age-appropriate qualitative data collection methods implemented by skilled enumerators, this will not generate quantitative data. Thus the team may not be able to report on either of the suggested indicators.

35. Bohl et al. 2018.

SMART



MEASURABILITY: GOOD PRACTICES AND TIPS

Good practice: When a direct measure is not feasible, choose a proxy measure of the change you want to see.

In our projects, we aim to have direct or “coincidental” indicators that coincide with changes specified in the objective statements. However, due to limited resources; project timeframe; or ethical, protection or other constraints, sometimes there is no choice but to select a proxy measure, either “leading” (preceding the expected change) or “lagging” (following the expected change). For example, measuring the percentage of people willing to enter into an economic relationship with other ethnicities may be considered a proxy, leading indicator for improved inter-ethnic collaboration.



Example 6 from the field:

Leading, coincidental and lagging indicators for education project

An education project aimed at *Increasing access to six years of quality primary education for boys and girls* (SO) focuses on addressing two main barriers identified during the assessment: (1) lack of schools in the target communities as the main barrier to enrollment, and (2) parents’ attitudes to girls’ education as the main barrier to keeping girls in school. The two IRs focused on these two main barriers, and could have had the following choices of indicators:

IR1: Increased enrollment of boys and girls		
Proxy, leading indicator: a. Number of new schools established	Direct, coincidental indicator: b. Number of students enrolled in the new schools	Lagging indicator: c. n/a
<i>The assumption being that establishing new schools will lead to increased enrollment and access.</i>	<i>Directly measures changes in access to education in target communities.</i>	<i>A lagging indicator would be hard to identify here as it would take us to the achievements beyond what this IR refers to (i.e. graduating from schools, example f below)</i>
IR2: Parents improve attitudes toward keeping girls in school		
Proxy, leading indicator: d. Percentage of parents reporting that they intend to keep girls in school to complete their education	Direct, coincidental indicator: e. Dropout rate per school semester, disaggregated by gender	Lagging indicator: f. Number of students graduating from primary school, disaggregated by gender
<i>The parents report their intent, but we don't know whether this would indeed happen.</i>	<i>The indicator may be hard to track, e.g., school may not provide records.</i>	<i>This should be measured after six years of primary education. Will the project still be running? Is this going to be captured at the SO level?</i>

S M A R T

Proxy indicators are often the best choice for higher-level objective statements/SOs.

When and where to use what type of indicator

Proxy indicators are often the best choice. They can be developed for any objective statements; however, in practice, they are typically used at higher levels of the RF, e.g., SOs. Lower levels, such as outputs or IRs, are much more within our management control, and therefore easier to develop direct or coincidental indicators for. As in the example of IR2 f, a lagging indicator at this level often points to a higher-level achievement that should either be measured at the SO level or perhaps is a proxy for an SO. When considering proxy indicators, refer back to your barriers identified in the assessment to narrow down the focus of the indicator and its definition. Finally, be careful when considering lagging indicators. Will your project last long enough to capture this information?

Good practice: When foreseeing challenges with a data source or data collection method, opt for the next best alternative.

When an output-level indicator measuring knowledge change is not feasible or data collection is too onerous, consider using simpler, low-effort indicators at the output level.

When determining potential alternative data sources and data collection methods for a specific indicator, go back to the description of change captured in the indicator and consider whether such change is observable or must be reported. If observable, consider who is best positioned to notice this change and, if it needs to be reported, who else besides direct project participants could be reliable reporters. For example, for information on children's physical or emotional changes brought about by participation in the project, the parents or caregivers may be a much more reliable and more accessible data source than the children themselves. Similarly, the observations of interactions among children by animators or teachers may be sufficient to capture change in the social or interpersonal dimension of well-being.

Strategically balance the level of effort, resources and rigor required to ensure accountability to the donor and enable evidence-based adaptive management. When an output-level indicator measuring knowledge change is not feasible or data collection is too onerous, consider using simpler, low-effort indicators at the output level (e.g., number of training participants, number of training sessions) and get the information about the results of these efforts through light monitoring. This will help you get early insights into the progress of the activity—including training content, delivery method and knowledge retention—thus enabling the team to make timely adjustments, if needed, and focus more rigorous monitoring efforts on measuring and capturing changes at higher levels (e.g., changes in practices introduced through new knowledge, changes at the SO level resulting from adoption of new practices, etc.).

If, however, a donor requires reporting on lower-level indicators, such as knowledge changes, and the project's scope and timeframe allows for an endline survey, consider including simple knowledge check questions in the endline survey (and baseline and midterm, as appropriate). This data, along with higher-level questions on practices, will enable analysis and documentation of the extent to which changes in knowledge lead to changes in practices. It is recommended that light monitoring is used for project adaptive management needs. This approach would reduce the frequency of data collection and lower the level of effort required for rigorous reporting on knowledge change, but you will have good enough evidence from light monitoring for adaptive management purposes.

S M A R T



Example 7 from the field: Light monitoring in practice

Data collection for light monitoring was initiated immediately after training activities began, but was stopped once it indicated that they were effective.

An agricultural project aimed to improve the livestock-based livelihoods of poor herding families (SO) by helping them to adopt improved livestock management practices (IR). Project activities comprised mostly trainings aimed at increasing male and female herders' knowledge (output) about selected livestock shelter practices and feeding practices, identified as critical, and not already in use at assessment time. The team was unsure how to conduct a knowledge check with an illiterate population, as a typical test-based approach was not feasible.

The team decided to track the number of participants in training sessions (typically considered to be at the activity level) as part of the formal MEAL plan for the donor and use **light monitoring** to check the retention of key messages immediately after the training using individual interviews with a small, purposeful sample of herders. This approach gave the team early insights into the effectiveness of the training content and delivery, as knowledge represents a critical precondition for progressing to the next level of change, i.e., the adoption of the new practices. In this example, data collection for light monitoring was initiated immediately after the training activities began, but was stopped once it indicated that they were effective. The more rigorous data collection effort (and resources) focused on the IR-level indicator measuring the percentage of herders adopting selected practices.

This is a good approach to balancing information needs for project management with available resources and rigor requirements.

SMART



ACHIEVABILITY: TYPICAL CHALLENGES AND IMPLICATIONS

Challenge: Indicators may not be achievable given the nature of change

The achievability challenges may be two-fold: (1) stemming from *what* is chosen as an indicator or (2) arising from *how much* change we anticipate happening, as expressed in the targets.

If a proxy indicator is a lagging indicator, our project may not last long enough to capture that change.

Challenges with “what” typically occur in complex, multi-year projects, where a change in terms of quality and quantity may be hard to describe and predict, or challenging to track. In these cases, we often follow the best practice of using proxy indicators. However, if a proxy indicator is a lagging indicator, our project may not last long enough to capture that change. This challenge can also happen in shorter projects that are tied to the seasonality of agriculture or school years, when project duration and its activity dynamics do not match the seasonality or school calendar, hence the change measured with the indicator, while feasible, will not be captured within the project timeframe.

The “how much” challenge can occur in any project and relates to how realistic our targets are. As with measurability challenges, the implications are severe as we will be unable to measure whether the intended change has taken place and ultimately will not have anything to report as an achievement for the specific indicator.



Example 8 from the field: Seasonal effect on attainability

An 18-month project implemented across five regions aimed to support vulnerable households to meet their food needs during the lean season by restoring their agricultural production of potatoes as a key staple and cash crop. The two IRs focused on improving agricultural practices for growing potatoes (IR1), and improving practices for potato storage (IR2). The project’s SO and proposed indicator were:

Strategic objective	Proposed indicator
Drought-affected households meet their essential food needs throughout the year.	Average reduced Coping Strategy Index (rCSI) (to be collected on a random sample of targeted households at the end of the lean season)

At first sight, this indicator seems very good. It uses a standard, tested food security measure, which—through questions exploring the five most common behavioral changes in response to food shortages—adequately reflects the food security status of a household. However, the start date of the project and its duration may have a significant, potentially detrimental impact on whether this indicator will be feasible to measure the achievement of the SO.

S M A R T

Depending on the start date, the project's duration may include one or two lean seasons. If there is only one, we may be unable to show any change unless we have secondary data that can serve as a baseline. Also, the project may need time to train all the targeted households and accompany them to start adopting new practices, which means full adoption may not be evidenced before the second potato planting/storage season, which in turn means that the induced changes in negative coping behaviors during the lean season may not be well aligned with proposed baseline/endline data collection points.



ACHIEVABILITY: GOOD PRACTICES AND TIPS

Good practice: Choose the most appropriate indicators to capture progress/change during the project's lifespan.

Add internal proxy indicators when the proposed intervention will not last long enough to capture the changes reflected in the donor-mandated indicators.

This often requires a delicate balance between what is ideal versus what is possible to measure, to enable adaptive project management and reporting on progress, and to capture learning. If you realize a proposed indicator is not meeting this dimension of the SMART criteria, you need to go back to the TOC (Step 1) and try to better describe how change or success should look *by the end* of your intervention. The indicator may need to focus on initial signs of a longer-term change you are aiming to introduce through the intervention, if these are expected to occur and can be measured within the project timeframe. It is also good practice to add internal proxy indicators when the proposed intervention will not last long enough to capture the changes reflected in the donor-mandated indicators—such as rCSI in the example above—or if these indicators will not generate useful data for project management and learning.

Good practice: Consider whether an alternative analysis method can account for seasonality challenges.

The doer/non-doer comparison, discussed under [Use of disaggregated data](#) on Page 24, may be a solution to mitigate challenges arising from the seasonality effect on baseline-endline comparisons. For a doer/non-doer analysis to work, the project has to have a number of participants who are doers (e.g., attend training, adopt new practices) and a number of participants who are non-doers, to enable a comparison between the two subgroups and hopefully identify differences that would testify to the success of the project's approach and results. The assessment will again be an important source of information, as will the results of the light monitoring, as both should point to the likelihood of having doers and non-doers, and the feasibility of such an analysis.

S M A R T

Good practice: Be sure to set reasonable targets for all your indicators.

Target setting is a good test of the feasibility of indicators.

Without targets, we would not know how much change we expect to see reflected in the indicators. Targets define whether there will be an expected increase or decrease, and by what magnitude.³⁶ Target setting is a good test of the feasibility of indicators; if you can't set the target, your indicators may need to change.

Targets should be ambitious yet achievable, in line with available resources, the project timeframe and dynamics, and the operating environment. Target setting is informed by:

- Relevant industry or professional standards (for example, Sphere lists minimum requirements in humanitarian response, including for shelter, food security and nutrition, and water supply, etc.).
- Needs assessment findings and/or baseline values.
- Results of past similar projects, either in the targeted country or another country with similar context.
- Project scope and budget (how much change can you anticipate given project resource investment) and activities schedule (when can you reasonably expect change?)

A 100% target is a likely sign that there is a problem with the indicator.

A 100% target is a likely sign that there is a problem with the indicator. The indicator may be too easy or too vague and therefore not be very useful for project management. Review in particular the Specificity and Time dimensions of the indicator. For example, at the output level, adding a time dimension may be a way to make the indicator more useful for project management, e.g., instead of measuring *percentage of NFI winterization vouchers redeemed* (which is likely to be 100%), consider *percentage of NFI winterization vouchers redeemed before the onset of the winter or by December 1*; this is the date when vouchers should be redeemed for the higher-level results to occur (that vulnerable people are protected from the cold), so tracking *when* the vouchers are redeemed is in fact what is most useful for project management.

“Determining success first requires identifying what changes are needed, which requires asking, ‘what is the continuum and where are the stakeholders on the continuum?’ Only after answering these questions can we say how much change needs to take place to be considered a success.” Church and Rogers (2006)

36. USAID 2010c.

S M A R T

Tips for target setting

Start from the outputs and work your way up. Targets at the output level are directly linked to the budget and activities schedule: we know how much of what is supposed to happen when. Our management control decreases as we move up the objectives hierarchy, but findings from needs assessments, previous projects, and past results, etc. should help us estimate how targets at the higher levels of the RF/Proframe should evolve.

The targets should be set in reference to baseline data values. Without this information, there is a risk of setting unrealistic targets or even setting targets that are too easily or even already achieved. A baseline is often required for higher-level indicators so you can show the difference—hopefully an improvement—between before and after the project started. Be sure to include a plan for baseline data collection in the project activity plan at the time of submission and then in the detailed implementation plan (DIP) after the project is approved. If baseline data is not known at the time of target setting, most donors accept a note saying it will be determined at baseline.

Target setting is a good test of the feasibility of indicators.

If the donor requires that the targets be set at the time of proposal submission, use the best estimates based on assessment findings or past projects, and state that targets may be revised once baseline data is obtained. In all cases, specify when the baseline data will be collected. Always use the baseline to check feasibility and revisit targets for all indicators. After the baseline is collected, you may be much better positioned to articulate more realistic expectations. Target setting needs to **“mirror” the unit of measure used in the indicator.** If the indicator is a number AND a percentage, set targets in both the number and percentage.³⁷

Set targets per disaggregation category and over time. Few donors require targets to be submitted with this level of detail, but it is strongly recommended that this is done in internal design documents. This level of detail helps harmonize what is proposed in the DIP with what is expected as a result, and ensures that due attention is paid to each subgroup targeted by the project. The exercise of comparing actual achievements with targets for each disaggregation category over time can reveal hidden assumptions related to performance of a specific subgroup in a specific project performance period. Note that while donors may not request disaggregated targets at the time of project submission, they often do request reporting per disaggregation category.

37. Having indicators expressed as a number AND as a percentage is generally not considered good practice, but may be required by some donors, e.g., BPRM.

S M A R T

Don't be afraid to propose adjustments to project targets, especially if changes in the context or activities mean that the initial targets are not relevant or will not be useful.

Consult with partners and other stakeholders to ensure your targets are as realistic as possible. Overly ambitious targets can make even a successful project look like a failure if they are not met. Review the targets regularly, e.g., during annual project review meetings, as well as after the baseline data collection and midterm evaluation. Don't be afraid to propose adjustments to project targets, especially if changes in the context or activities mean that the initial targets are not relevant or will not be useful. But don't adjust targets every quarterly meeting and/or just because project results are slightly over or above plans; the indicators and targets should be used to generate a discussion within project teams about why results are higher or lower than expected and whether some adjustments to project activities or approaches are warranted. In all cases, changes should be appropriately justified and communicated to the donor and other relevant project stakeholders.

S M A R T



RELEVANCE: TYPICAL CHALLENGES AND IMPLICATIONS

Challenge: Indicators are not relevant

This typically happens when an indicator is misplaced in the Proframe, i.e., it measures either a higher- or lower-level objective statement. It is measuring not the core but a small secondary aspect of the objective statement that is not useful for learning about the TOC, or when the indicator is poorly worded and does not result in useful information (e.g., *percentage increase in knowledge* versus *percentage of people with x knowledge*). When this happens, a MEAL system developed using these indicators results in irrelevant information that does not capture project results or key aspects of the TOC and is therefore useless for project management and learning.



RELEVANCE: GOOD PRACTICES AND TIPS

Good practice: (Re)consider how the indicator refers to the level of change stated in the objective it measures.

When facing a relevance challenge, refer to the guidance and tips presented in Step 3 of the process for developing indicators described above. Much of literature also advises thinking about the **indicator type**—whether it is an output, outcome or impact.³⁸ Some donors, such as BPRM, require indicators to be classified according to these types in the proposal submission. Simply put, the indicators need to track or match the different levels of change they are measuring. The **output indicators** track the deliverables expressed as outputs, the immediate benefits of the activities, such as gaining new knowledge, accessing a service or receiving support (e.g., *number of entrepreneurs trained in accounting*).³⁹ **Outcome indicators** measure the expected short- or medium-term changes expressed in the IRs (e.g., *percentage of trained entrepreneurs applying new accounting knowledge*), and end-of-project results captured in the SOs (e.g., *percentage of entrepreneurs reporting increased profit*). When projects include **impact indicators** that correspond to the project goal, these are very few in number, usually only one (e.g., *percentage of entrepreneurs who report improved quality of life*). Note that most donors do not require indicators for the goal,⁴⁰ and nor does CRS.

38. Note that in addition to output, outcome and impact indicators, there are also input indicators representing an organization's investment or input into a specific activity measure—labor, resources, time or materials, e.g., number of training sessions organized or number of NFI kits purchased. These are typically not developed, as activities and inputs are best tracked through the DIP and budget rather than a specific, explicitly stated indicator.

39. People in Need 2021.

40. CRS and Humentum 2019.

S M A R T

Tip: Using indicator types

While using indicator types sounds appealing, in practice there is no straightforward rule as the type of indicator originates from its association with Proframe statements, and how ambitious the outputs, IRs and SOs are. An output indicator in one project may be an outcome in the other. For example, in emergency projects, *number of people reached by intervention* is often placed at higher levels of the RF, while in development projects, *number of people reached by a specific activity* usually measures a lower-level objective.

Good practice: Use a peer or external party review to ensure the relevance of your indicators.

In most cases, a review of the objectives and indicators—typically by an external party who has not been involved in the design—will highlight irrelevance issues, e.g., it is easy to recognize an output-level indicator misplaced to measure an IR or even an SO, or an impact-level indicator associated with an SO or IR. If this is the case, something is wrong and the team should revisit Step 3.

Good practice: Be sure that the complement of selected indicators together helps manage the project and gauge its results and learning about the TOC.

While the importance of having a strong link between indicators and the objective statements they are supposed to measure cannot be understated, their relevance goes beyond attachment to the specific objective statement and needs to reference the TOC. When checking relevance, it is essential to step back, look at the TOC again and review the entire set of indicators—the performance monitoring plan as a whole—to check whether selected measures *together* will indeed ensure effective performance management and generate information that is relevant, timely and useful for adaptive project management and learning. This is another opportunity to verify disaggregation categories and analysis plans for the entire set of indicators.



Example 9 from the field: Education project

A five-year project aimed at improving educational outcomes for girls, so they were empowered to participate equally with boys in the sustainable development of their communities (goal). Since the donor requested goal-level indicators⁴¹ the team included: *number of students who are continuing education in Ministry of Education (MOE) schools, disaggregated by gender and location* and *number of classes handed over to the MOE*. The project had a three-pronged approach with the following SOs and indicators:

Strategic objectives	Indicators
(1) Girls have improved equitable access to quality primary education that is age-appropriate, inclusive and gender-sensitive.	Number of students enrolled, disaggregated by gender and location.
(2) Education for girls for targeted provinces is sustained.	Number of community-based classes that receive MOE commitment for continued support, disaggregated by location.
(3) Female teachers access Ministry of Education (MOE) employment opportunities as credentialed teachers (formally recognized by the MOE).	Number of teachers certified by the MOE, disaggregated by gender.

The selected indicators at the SO level were well chosen based on their type and relevance to the objectives. However, a repeated review of the TOC describing progression from SOs to goal, and a “big-picture” review of the proposed indicators revealed potential gaps. The team had not proposed tracking regular attendance, drop-outs, and the progression of children through classes (completion of grade 1, enrollment into grade 2, etc.), all of which are essential measures to ensure relevance of indicators and tracking of “unknowns” in the TOC.

41. In this case, the donor requested indicators at the goal level. The team thought that these indicators could be captured over the five-year span with the gradual transition of community-based education (CBE) schools and students to the MOE jurisdiction.

S M A R T



TIME: TYPICAL CHALLENGES AND IMPLICATIONS

Challenge: Indicators lack a time dimension

Challenges here relate to the absence of information on relevant timeframes for the targets, the appropriateness of data collection frequency for a specific indicator and, in some cases, a time qualifier within the indicator description itself, if an indicator measures a change that entails a time-sensitive phenomenon. Failure to properly include time elements in the indicator often results in diminished indicator quality, indicators that are too easily reached, and indicators that are repetitive across different levels of the project hierarchy or irrelevant to the project and its planned activity dynamics, and thus useless for project management.



TIME: GOOD PRACTICES AND TIPS

Good practice: Add the time qualifier to indicators that require it.

If an indicator measures a time-sensitive change, the best approach is to add the time qualifier within the indicator wording. A time qualifier is particularly important at the output level and for qualifying 100% targets on outputs that are largely within our management control. For example, the indicator mentioned above, *percentage of target households who receive winterization NFI kits before the onset of the winter*, is very different from the one without the time qualifier. Without a time qualifier, it is likely that 100% of participants will receive kits; however, perhaps only 90% will receive them in time for the cold weather due to multiple factors. In this situation, the wording “*before the onset of winter*” is probably best for documents shared with the donor, although we would want to set an internal deadline for ourselves of before the onset of winter (e.g., November 15 or December 1). In so doing, we could discover whether our distribution plans meant that many households would receive the kits only after the target date (i.e., the procurement process may take more time), and put corrective action in place (i.e., organize more distribution teams to reach more households faster once procurement was completed). This indicator would be tracked weekly to enable timely adjustments to activities, schedule, etc., even if reporting to the donor only on a monthly or quarterly basis. The time dimension of the indicator may come from the context or the project activity plan/DIP and TOC. In most cases, a certain sequencing of activities delivers the specific pathway of change, e.g., delivery of seed inputs or knowledge on planting techniques needs to happen before the planting season. The time dimension informed by the context is most critical for emergency or seasonally tied projects where timing of support is critical.

If an indicator measures a time-sensitive change, add the time qualifier within the indicator wording.

S M A R T

Good practice: Align the data collection frequency of your indicator with the project's activity plan/DIP.

The data collection frequency should reflect the estimated pace of change, and what is useful for project management.

Some indicators may require data collection from project start to project end, but most will not. The data collection frequency should reflect the estimated pace of change, and what is useful for project management. In addition to the data collection frequency (daily, monthly, quarterly, semi-annual, etc.), indicate when data collection should start and end. This can be a specific calendar month in which data collection should occur based on environmental and seasonal considerations, or it can be a project month or quarter, based on the activities schedule and the TOC, i.e., assumptions of when this specific level of change will start to happen. This is helpful for project planning and budgeting purposes. It also checks the feasibility of capturing a change with a proposed indicator by following the data collection frequency suggested in the plan. Also, given the type of change captured at each level of the ProFrame, the data for indicators at lower levels of the objectives' hierarchy (outputs and IRs) need to be collected more frequently than data for indicators at the SO level. When data collection frequency at the output or IR level is higher than at the SO level, something is wrong. The change at the SO level needs to happen as a result of a collection of activities/ outputs and the delivery of the IRs, therefore it cannot occur faster or more frequently than lower-level changes.

Good practice: Set the targets per period.

Even if you need to present the targets to the donor as cumulative, end-of project (EOP) targets, it is useful to break them down according to the activities schedule, to ensure your data collection frequency corresponds with specific project benchmarks and that the EOP targets are indeed realistic. The length of the period—whether it is monthly, quarterly, semi-annual or annual—depends on the duration of the project and key project activities, specifically the time it takes to deliver outputs, a change at the IR level and then the EOP result at the SO level. When setting targets per period, be careful when specifying EOP targets. For indicators expressed as numbers, it is appropriate to have the cumulative value of all project quarters or all years as the EOP target: the sum of targets of all periods. For indicators expressed as a percentage, a cumulative target is not appropriate as a percentage should not be summed up or averaged. The EOP target for indicators expressed as a percentage often equals the percentage target for the last period, although this highly depends on the project DIP and the activities planned for the last period of project implementation.

S M A R T**Tip: Set targets per cohort**

If your project is implemented through sequential cohorts of project participants, consider setting targets per cohort, and then see how these align with project years or the donor reporting calendar. If the targets are set only per calendar year, and cohort dynamics do not align, you may not be able to calculate your indicators and achieve your targets, especially for the last cohort for multi-year projects.

Indicator performance tracking table (IPTT)

A useful tool for setting targets is the **indicator performance tracking table (IPTT)**, in which targets for each indicator are specified for each relevant project period (usually quarters). Although mainly used as an implementation supporting tool,⁴² documenting actuals and reporting on progress against each indicator's target, experience has shown that the IPTT helps facilitate more granular and detailed thinking about realistic targets over time, especially when a project involves multiple partners.⁴³

42. Note that BHA requires an Indicator Tracking Table, which, at the time of submission, resembles the structure of the PMP, and, at the time of reporting, an IPTT (for more information, see: [USAID-BHA ITT suggested template](#)).

43. For more information on the IPTT template and how to fill it out, refer to [A guide to the MEAL DPro](#) (CRS and Humentum 2019).

Annex 1: Example—Writing indicators to include direction of change and target

<p>SO: Vulnerable internally displaced and returnee households have sustainable livelihoods that meet their basic needs.</p>	
Indicator typically used	Indicator with direction of change and target
<p>Percentage of targeted households that report that they were able to meet their household needs over the previous month, disaggregated by status (IDP versus returnee).</p>	<p>[X percentage point] increase in the percentage of target households that report that they were able to meet their household needs over the previous month, disaggregated by status (IDP versus returnee) the end of the project [specify date].</p>
<p>The baseline plans, target and period in which the target should be achieved are specified elsewhere, most often in the MEAL plan or other tool that supports and communicates details of indicators' design.</p>	<p>Very few donors request indicators that specify a direction of change (“increase”) and targets (X% percentage point change in comparison with baseline) within the indicator, but, when encountering these requirements, present the indicator in this way.</p>

Annex 2: Frequently asked questions about qualitative indicators

What is the difference between quantitative and qualitative indicators?

Quantitative indicators help us understand **how much** of something is happening, while we know or hypothesize from the TOC **what** that something is. With qualitative indicators we investigate the **what** as we are not sure what to expect; the indicator will help us find this out. Indicators such as *the percentage of women reporting a high sense of agency*, or *the percentage of caretakers reporting an improved sense of the well-being of their children* are considered quantitative in this guidance since the “what” within the indicator—sense of agency, well-being—has been clearly defined and then measured using quantitative methods and tools. The quality has been “quantified,” measured through closed-ended questions that generate quantitative data, which are then analyzed using statistical methods. Quantitative indicators require quantitative data collection methods. They collect data through structured data collection tools containing closed-ended questions (e.g., yes/no questions, multiple-choice questions, rating scales, etc.) and result in a number, percentage, ratio or rate. Qualitative indicators require qualitative data collection methods such as FGDs, participatory ranking exercises and semi-structured interviews. They generate qualitative data—a narrative description of what people do, think or believe.

What could a qualitative indicator tell us, and how reliable and valid is it?

A qualitative indicator provides insights into a change experienced by participants; however, it cannot tell us how much of that change has happened or the proportion of the target population that experienced that change. That is the job of a *quantitative* indicator. Qualitative indicators reveal participants’ perspectives and opinions about the project, either positive or negative, which the project either intended or did not intend to achieve. Although qualitative data, and therefore qualitative indicators, have limited generalizability potential and, given the subjectivity of qualitative data, are often seen as unreliable and prone to bias, they can still provide valuable insights into participants’ opinions, judgments and perceptions about changes the project has or has not made. To ensure the greater reliability and validity of a qualitative indicator and the associated qualitative data, it is recommended that you:

- Ensure triangulation by involving two or more data sources, and using two or more qualitative data collection methods.⁴⁴
- Ensure good documentation of the data collection process, including interview scripts, focus group discussion guides, etc., so the process can be replicated using the same approach.
- Engage several team members to do the data coding and analysis to minimize the potential bias that could occur during analysis of qualitative data.
- Ensure consistent analysis methods through all data collection rounds including a consistent codebook.

44. Dzino-Silajdzic 2018.

Triangulation helps increase the validity of the results, while procedures that ensure consistent data collection and analysis ensure the process can be replicated and thus result in more “stable” and reliable results.⁴⁵

Do we need to have both quantitative and qualitative indicators?

No! At this point, we do not need to have both qualitative and quantitative indicators. We do need *qualitative data* though, but this is not the same as a qualitative indicator. In a typical project MEAL system, we often use qualitative data to explain the “why” and “how” behind quantitative indicator results. This practice is strongly recommended (see below), especially for quantitative indicators measuring SOs. In addition to providing a more in-depth explanation of quantitative achievements, collecting qualitative data after analysis of SO-level survey results ensures community input into the interpretation of quantitative results, thus also supporting our accountability to project participants.

But what do donors say? Do they require both quantitative and qualitative indicators?

No! Very few donors require qualitative indicators expressed in descriptive terms. Most rely on quantitative measures of project performance and consider quantitative indicators with qualitative dimensions sufficient to meet project stakeholders’ information needs and capture project results. Qualitative indicators should be used only when the donor requires them, and/or when the project team is not certain about the results of their intervention and cannot fully specify the expected change. But donors are increasingly interested in qualitative data. For example, in the Resilience Food Security Activities technical guidance,⁴⁶ USAID asks for “qualitative inquiries” that may be in the form of a:

- Qualitative study: A stand-alone qualitative data collection effort, typically exploratory in nature to better understand dynamics of phenomena related to higher-level results, e.g., a qualitative study to inform or refine the technical approach; or
- Qualitative monitoring: A regular qualitative data collection to capture less tangible results related to behavior change, outcomes or context monitoring including conflict dynamics, unintended consequences, secondary adoptions of promoted behaviors, etc.

Each needs to be planned at the time of project submission using Qualitative Inquiry Planning Sheets (QUIPS)⁴⁷ listing the purpose, inquiry questions, type of data or level of change the Proframe qualitative inquiry intends to explore; the data collection methodology (data sources, sampling, tools, implementation plan and timeline, frequency and timing, any training requirements, data management and quality assurance); and the analysis plan including key disaggregation groups, deliverables, utilization/application, limitations/risk and ethical considerations.

45. Cypress 2017.

46. USAID 2020.

47. For QUIPS template, prefer to Annex V of [Technical guidance for monitoring, evaluation and reporting for resilience food security activities V2.0](#) (USAID 2020).

If we follow donor guidance and use only quantitative indicators, we should be ok, right?

No! Even when a donor requires only quantitative indicators, they often require interpretation of the quantitative results—the why and how behind them—which is best obtained through collecting qualitative data. Quantitative indicators are expressed as a number, percentage, ratio or rate that represents an achievement. If, when compared to the target, the achieved value signals a problem (e.g., results are below expected targets or are lower for a specific subgroup, etc.), the best practice is to use qualitative data collection methods and tools to learn why and how the surprising results occurred. Qualitative data collection methods are used to probe reasons for certain groups being more successful than expected or more successful than other subgroups, so the project teams can learn about strategies of success for potential replication. Qualitative data is often needed to provide a detailed account of issues that are less tangible, and require nuance and contextual details. With **quantitative indicators, we learn *what is happening***, while **qualitative data helps us understand *why and how it has happened***. This information is critical to informing adaptive management actions, but this is not a separate indicator: it is additional (qualitative) data that helps add richness to the analysis and interpretation of quantitative indicator results. Typical occasions when qualitative data is most useful are:

- To provide early insights on whether quantitative indicators are likely to be achieved and to enable adaptive management. This type of monitoring is called light monitoring⁴⁸ and it is typically done for a limited time, early in project implementation and for selected activities when we are less sure how things will go. Data generated through light monitoring may be both quantitative and qualitative (e.g., using selected questions from a baseline survey with the addition of open-ended questions), but due to a small sample, the findings cannot be generalized and are treated as qualitative data.
- To probe and interpret quantitative data. This is typically done after administering surveys or other quantitative methods. In these cases, the qualitative data is critical to understanding why or how certain results have been obtained on any of the quantitative indicators included in the Proframe.

What are some examples of qualitative indicators?

Here are some examples of qualitative indicators:

- The nature of interactions between refugee and host community youth.
- Challenges experienced by girls in accessing formal primary education.
- Community perceptions of the “other” ethnicity.
- Young women’s visions for their roles in their household or community.

48. For more information on light monitoring, consult [Practical guide: Monitoring for problem-solving, adaptive management, reporting and learning](#) (Morel et al. 2019).

The key to phrasing qualitative indicators is to ensure they are as neutral as possible. They should express the phenomena being measured rather than the expected or desired situation. For example, *positive interactions between refugee and host community youth* would point to a desired situation that represents a target for qualitative indicators.

Quantitative indicators should be SMART, but what about qualitative indicators? Should they also be SMART?

Qualitative indicators should be SPICED. The SPICED framework was developed to help teams collaborate more effectively with communities to develop qualitative indicators.⁴⁹ The SPICED concept says that indicators developed collaboratively are stronger when they are: Subjective, Participatory, Interpreted, Cross-checked, Empowering and Diverse.⁵⁰

SPICED checklist	
Subjective	Recognize that target groups have a special position or experience which gives them unique insights, opinions and perspectives that are important to the monitoring process.
Participatory	Develop indicators together with those best placed to assess them. This means involving a project's direct participants, or other relevant project stakeholders.
Interpreted	Ensure staff and partners can interpret the indicators and make meaningful use of the information generated from the data collection process. In other words, it should be easy to analyze the data coming out of the data collection process.
Cross-checked	Ensure it is possible to cross check the validity of the assessment by comparing different indicators and results or triangulating data collected through the indicator. Also, the conclusions that partners and program staff come to as a result of the monitoring process should be replicable by other people at a later stage.
Empowering	Ensure the process of formulating and assessing indicators is empowering in itself and allow target groups and other project participants to reflect critically on the changing situation as a result of the program intervention.
Disaggregated	Make a deliberate effort to seek out different perspectives 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 and compared over time.

49. Much of the literature suggests that the SPICED checklist should be used to check whether indicators are developed in a participatory way, whether or not they are quantitative or qualitative. See [CIVICUS Monitoring Toolkit](#).

50. [A guide to the MEAL DPro](#) (CRS and Humentum 2019) checklist presented here is a simplified version. The guide uses further qualifiers: Interpreted and Communicable, Cross-checked and Compared, Diverse and Disaggregated.

Here is an example of a qualitative indicator for a peacebuilding project, run through the SPICED checklist.

Indicator: Perception of “other” ethnicity	
Subjective	This indicator is subjective as it calls on identified data sources to share their perceptions (opinion, judgment) of an ethnicity other than their own.
Participatory	The perception of “other” will be examined through FGDs in a participatory way with project participants.
Interpreted	Staff and partners will engage in participatory data analysis (coding) and interpretation. Changes in perceptions of the “other” over time will be analyzed in relation to project activities and objectives.
Cross-checked	The information collected through this process will be compared with findings from other project indicators (at same or other levels of the ProFrame). The process of data collection and analysis will be documented for future replication and to enable comparison over time.
Empowering	The process of data collection gives voice to the project participants to reflect critically on their perception of ethnicities other than their own, and how these perceptions have changed over the project’s duration (changes may be both positive and negative).
Disaggregated	FGDs will be conducted separately with participants from different ethnicities and with men and women. (Other disaggregation categories, such as age, could also be relevant).

Annex 3: Using closed-ended questions to measure quantitative indicators⁵¹

TYPES OF CLOSED-ENDED QUESTIONS

Quantitative indicators are measured using closed-ended questions that are included in quantitative data collection tools. The simplest are **binary or two-response questions**, e.g., *Have you attended the training?* with yes/no response options.

Scale-based questions offer more response options, and therefore enable getting more nuanced feedback about the issue you want to explore. Scale-based questions typically indicate the respondent's level of satisfaction, agreement, etc. For example, an indicator measuring satisfaction with post-training support can be measured by the question: *How satisfied are you with the post-training support you have received?*, with response options ranging from “very dissatisfied” to “very satisfied.” The scale can also be used with statements. For example, the same indicator above can be measured through rating a specific statement, e.g., *I am satisfied with the support I have received after the training*, with response options ranging from “strongly agree” to “strongly disagree.”

These are also called **one-choice response questions** or statements, as respondents are allowed to check only one option as their response. A form of one-choice response question is often used in knowledge tests, with one of the response options being the correct one.

Another type of closed-ended question is a **multiple-choice question**, in which the respondent is asked to check all that apply to their specific situation. An example of such a question is: *Which type of support was most useful to you?*, followed by a list of options for the respondent to choose from, e.g., *face-to-face consultation with an expert, online expert available to respond to my questions, resources on specific subjects*. They are often used as follow-up questions to scale-based questions. The respondent can choose one, some or all applicable options.

So-called **scenario-based or situational questions with multiple-choice response options** are often used to situate a respondent in a specific scenario and check how they would act. These typically describe a specific situation and ask the respondent to check all that apply to their emotions, attitudes or behaviors related to that situation. For example, when checking *percentage of children who report being able apply problem-solving skills*, we may ask a respondent to recall the child's behavior in school during the preceding week, and check all that apply with the following possible options: *The child fought a lot, the child shared their toys with others, the child was nice to at least one school mate, and the child took things that were not theirs*. Scenario-based questions can also request only a single response.

51. These tips are aimed at meeting specific project management requirements rather than ensuring research rigor.

Depending on what your indicator measures and how it is defined, any of the closed-ended questions can work. Note that scenario-based questions elicit more details but require a lot of skill and work to develop, to ensure the options are appropriate, sufficiently exhaustive, and not overlapping. Scenario-based questions are often used with children or when the issue being explored is sensitive, and respondents are expected to be more prone to bias when using general questions with scales.

BASIC TIPS TO MAKE THE MOST OF CLOSED-ENDED QUESTIONS THAT USE SCALES

Ensure your question/statement addresses ONE phenomenon.

Closed-ended questions used to generate data for quantitative indicators often measure complex attitudes, feelings or behaviors. However, if you combine two or more issues in one question or statement, the response will not be accurate. An indication of this is the word “and” in your statement or question. For example, if the question is *How satisfied are you with the training content AND logistics?* the respondent may rate only one or perhaps average the response across the two, which in either case would not meet your requirements.

Choose sufficiently nuanced scales that correspond with your management needs.

The scale most frequently used is a Likert scale. This is a five- or seven-point scale, i.e., with 5 or 7 possible response options. The response options are usually from “strongly agree” to “strongly disagree.” Likert scales include a midpoint that indicates the respondent neither agrees nor disagrees and is thus neutral on the subject explored. In some cases, it is useful to go beyond a 7-point scale and add more nuance to responses (see [Annex 7](#) for an example). Which type of scale you use depends on:

- Whether the issue under investigation is sensitive. Some studies suggest that more options help ensure more nuanced opinions and enable better distinguishing of extreme opinions.⁵²
- The data source, in particular the age of respondents. A 3-point scale with simple agree – neutral – disagree options may be more appropriate for children.⁵³
- Your management needs and subsequent analysis plans. If you, as a project manager, are interested only in positive versus negative opinions, and/or the analysis plan suggests combining all positive response options into one category, it is very likely that you don’t need a nuanced scale. A simplified scale would meet your needs.

52. Bradburn et al. 2004.

53. Bohl et al. 2018.

Choose an odd or even scale depending on the issue under investigation.

In most of our work, collecting data that sides with a positive or negative option on the continuum is not the primary intention, so having a neutral midpoint is more appropriate. Even if the majority of responses center around the neutral option, this information is still considered insightful and useful for project management. Exceptions are data collection efforts exploring less sensitive issues (e.g., preference for location of a child-friendly space, or favorite snacks). In these cases, we seek data that will clearly side with either the positive or negative option.

Having a neutral option reduces the stress related to the respondent having to choose either the positive or the negative option. Experience has shown that, in self-administered surveys, such pressure may result in questions not being answered. Several studies point to potential bias associated with an even scale as respondents who may be genuinely neutral are forced to instead choose a positive or negative answer.⁵⁴

Ensure your scale/choice of responses is balanced.

If your scale has more positive than negative response options, you are introducing bias into the design of the questionnaire. This may be especially detrimental if the issue under exploration is sensitive, as personal bias may already predispose the respondent to avoid checking a negative response option.

Unbalanced responses

Unbalanced responses are often present in scenario-based questions, where it is often harder to develop a balanced set of options or some of the choices are too extreme to be selected by the respondent. For example, the team that worked on exploring child wellbeing had the following question in the data collection tool: “how does your child behave when someone takes their favorite toy?” The response options included: (a) seeks for help of the adult; (b) calmly asks the person who took it to return it back; (c) throws a tantrum; (d) withdraws in the corner; (e) does not care; (f) starts crying; (g) starts fighting with the person who took it. In this case, the team tried to offer an exhaustive list of options, however, only two are positive/desirable, while the remaining six potentially indicate problematic behavior. Some undesirable choices are very similar or are simply unlikely to be checked by parents due to a strong social bias. If you are uncertain that you exhausted all options, add “other (please specify)”. The key to ensure the response options are relevant and appropriate is field testing of the tool.

54. Bradburn et al. 2004.

Provide labels or descriptions of your scale options.

Scales that contain only numbers may result in inaccuracy. Rating options may have different meanings to different respondents (e.g., a score of 2 out of 5 may be considered “failing” in one context, but “nearly passing” in another). Also, when providing descriptive categories, make sure there is a clear difference between options. For example, “slightly agree” and “somewhat agree” have very much the same meaning. When using multiple-response questions, be sure to:

- Have an equal number of positive/desirable and negative/undesirable response categories.
- Start from the extremes, e.g., “extremely” and “not at all,” set the “mid-point” to represent neutrality and then use clear terms to distinguish ratings, e.g., “very”, “slightly.”⁵⁵ Include a neutral option to avoid exerting pressure on the respondent to side with either positive or negative opinions.
- Avoid having too many choices.

Remember, field-testing is the best way to ensure you have the appropriate response options in your final questionnaire.

Finally, remember that the more complex the scale is, the more complex the calculation of indicators will be.

55. [SurveyMonkey](#).

Annex 4: Step-by-Step Guidance to Develop Indicators

STEPS

STEP 1
REFLECT

STEP 2
RELATE

STEP 3
PRIORITIZE

STEP 4
REFINE

STEP 1: Reflect on the indicators using the draft theory of change

Who	Programming and MEAL team members, including partner staff and sector technical advisors.
Resources	Draft TOC, if available; if not, use RF or Proframe and their implicit underlying TOC.
How	<p>At the participatory session engaging the actors mentioned above, review the proposed pathways of change and draft TOC and discuss the following:</p> <ul style="list-style-type: none"> ■ How will we know we are successful in progressing through the pathways of change/TOC? How will success look? Are we expecting changes in a situation, a condition, a level of knowledge, an attitude or a behavior?⁵⁶ ■ Which parts of the TOC are we least certain about? Where do we have to make “leaps of faith” and describe the progression of change by adding explanations not captured in the RF?

At this point, don't worry about how the indicators are worded. This initial step is intended to identify potential signals of progress—i.e. types of change that the team expects to see over time—not to formulate fully fleshed out indicators. Discussing indicators early, based on the project's draft TOC, is likely to result in more meaningful indicators that are not just restatements of the objective statements but that better reflect the potentially complex pathway of change reflected in the project's TOC. Also, this discussion should help teams refine the overall project's TOC and the objective statements, because indicators often help clarify specific changes we want to see over time as a result of project activities.



Discussing indicators early, based on the project's draft TOC, is likely to result in more meaningful indicators.

TOC logic implies that over time we need to learn whether lower-level changes, i.e., outputs, are delivering the next level of change, i.e., IRs and then SOs. Lastly, focusing on the parts of the TOC we are least sure about—the “leaps of faith”, which in essence are our *programming assumptions*—will ensure that the MEAL system does not only help document results that teams are fairly confident of achieving, but will also include indicators that help the team monitor whether the changes that they are less sure about are taking place.

The indicators developed based on the draft TOC, or implicit TOC captured in RF/Proframe, should be useful for testing the TOC, rather than merely measuring objective statements in isolation from the if-then logical chain. Only when analyzed and reviewed together can indicators enable the testing of the if-then causality.

(Practical Guidance on Developing a Project's Theory of Change)

56. Church and Rogers 2006.

Having timely evidence that we are progressing (or not) over potential hurdles is essential for data-informed adaptive management.⁵⁷

STEPS

STEP 1
REFLECT

STEP 2
RELATE

STEP 3
PRIORITIZE

STEP 4
REFINE

Bringing in donor-mandated indicators during Step 1

Teams often like to bring in donor-mandated indicators during Step 1 and then fill in any gaps with custom indicators. Such an approach is possible, especially with experienced teams; however, it may result in over-emphasizing the inclusion of donor-mandated indicators rather than thinking through our own information needs for evidence-based project management. Ultimately, the project teams are responsible for delivering the higher-level outcomes, and donor-mandated indicators may not be sufficient to help teams manage for results.

57. Indicators measuring “leaps of faith” or programmatic assumptions are often subject to light monitoring, as the teams should check early on whether these assumptions are proving true. For more information on light monitoring, consult [Practical guide: Monitoring for problem-solving, adaptive management, reporting and learning](#) (Morel et al. 2019).

STEPS

STEP 1
REFLECTSTEP 2
RELATESTEP 3
PRIORITIZESTEP 4
REFINE**STEP 2: Relate the indicators with the types of change / objective statements in the Proframe**

Who	Programming and MEAL team members, including partner staff and sector technical advisors.
Resources	Draft TOC and draft objective statements for SO, IR and outputs. Any donor-mandated or standard sector indicators including agency global results indicators.
How	At the participatory session engaging the actors mentioned above: <ul style="list-style-type: none"> ■ Match the changes / draft indicators identified in the previous step with the project's draft objective statements. ■ Review donor-mandated indicators, agency global results indicators and other standard sector indicators and check for potential alignment with the changes / draft indicators identified in Step 1 for each level of the Proframe. <p>Check whether any objective statements have been missed. Don't worry about the indicator wording yet. Keep the indicators as detailed as is necessary to clearly understand what is being measured.</p>

The selected indicators should reflect the different levels of change captured in RF/TOC so they can signal that a change at each level is indeed taking place.

This step is necessary since each proposal submission requires presenting the indicators as they relate to different levels of change expressed in the objective statements. The objective statements capture change at different levels—SOs, IRs and outputs—while the TOC describes the if-then progression through these levels. The selected indicators should reflect the different levels of change captured in RF/TOC so they can signal that a change at each level is indeed taking place and the project is progressing along the pathway of change.

If some objective statements do not have corresponding indicators, or if the indicators initially identified based on the TOC do not have matching objective statements, something is wrong. This suggests that the TOC and RF may not fully align and that either refinement of the TOC or rephrasing of the objective(s) in the RF/Proframe may be needed. At this stage, be sure to include any mandatory indicators prescribed by the donor.

Additional sources of potential indicators in this step may come from previous project interventions, agency global results indicators and standard indicators (e.g., Sphere, GAIN) for the sector(s) your project focuses on. In all cases, consider whether you need to adjust these to the project context, and how much flexibility the donor allows (see [Annex 5](#)).



Frequently asked questions

Should we brainstorm to develop indicators?

No! While Steps 1 and 2 are popularly referred to as brainstorming, in practice there is very little brainstorming involved. This is an evidence-based process that rests on the results of the assessment conducted at the project design stage, chosen project strategies, and either an implicit or explicit TOC.

What should we do when we don't (yet) have a theory of change?

If you don't have a draft TOC to follow the process described in the steps above, use the project's preliminary RF and/or Proframe. Arrange objective statements in a means-to-end logic and probe possible assumptions and "leaps of faith" in progressing from one level to the next. This will help you articulate the project's TOC that underlies the RF/ProFrame.

Can Steps 1 and 2 be combined?

Yes! If you are developing the project's TOC and RF in parallel, or just using the RF or Proframe and implicit TOC to guide this process, then Steps 1 and 2 are typically combined. Note that in this case it is still recommended that you go through the initial questions in Step 1 before bringing in donor-mandated indicators. Donor-mandated indicators may not always be the best fit for project management purposes. Also, they often require adjustment that is best informed by a discussion of how success will look and other questions in Step 1. Finally, this approach will ensure you avoid the challenge of being too limited in describing changes resulting from the project as only those that donors state in their calls for proposals.

How are indicators associated with the different levels of objectives?

IR-level indicators typically relate to project strategy, measuring what the specific project stakeholders should be doing differently to deliver SO-level change. Output-level indicators typically reflect what these stakeholders need—knowledge, inputs, access to resources or services, etc.—to change what they do or how they do it. Output- and IR-level indicators need to be reasonably easy to measure so they can provide timely information for project management.⁵⁸ SO-level indicators reveal the central reason for the project. Indicators at this level should help us understand an end-of-project change: both how much change (quantity) and what kind of change (quality) we expect as a result of successful project implementation. Some donors require goal-level indicators, often population-based, meaning collecting data from census/entire population in the targeted country. If donors do not require them, it is recommended that indicators at the goal level are not included given challenges with attribution.

STEPS

STEP 1
REFLECT

STEP 2
RELATE

STEP 3
PRIORITIZE

STEP 4
REFINE

58. CRS 2015a.

STEPS

STEP 1
REFLECTSTEP 2
RELATESTEP 3
PRIORITIZESTEP 4
REFINE

Should we always include donor-mandated indicators?

Yes! Always include donor-mandated indicators. In so doing, we indicate alignment with donor priorities and objectives, as well as increasing our competitiveness during the application review by the donor. Because donor-mandated indicators must be aggregated across the organizations that received funding from a particular donor, there are often specific guidelines for definitions, sampling, data collection, analysis, etc. to ensure comparability. In many cases, donors have developed Performance Indicator Reference Sheets (PIRS) for these indicators; however, there may be some flexibility for contextualization and adjustment. Even when donor-mandated indicators may not be the best match, we are usually required to include them.

What if donor-mandated indicators do not match our TOC or objective statements?

Often it may not be a complete mismatch, but rather a challenge in terms of where to place the indicator in the Proframe and the potentially limited usefulness of the information collected through the donor-mandated indicator. For example, donor indicators may all correspond to the output- or SO-level change and may not be distributed across all levels of the Proframe. In this case, identify where donor-mandated indicators fit best, even if there are multiple donor indicators associated with the same objective statement and even if the fit is not perfect. Alongside these donor-mandated indicators, note the more specific and relevant signs of change identified in Step 1. This information may be useful for adjusting the donor-mandated indicator or as **a complement to the donor indicator**, either in a form of a custom indicator that better matches program management needs or as a contextualized definition of the donor-mandated indicator. Note that custom indicators that complement donor-mandated indicators do not need to be included in the formal submission to the donor; however, they should be included in internal project design and MEAL documents. Often these indicators are critical for adaptive management and intentional learning about the TOC introduced through the project.

How useful is the assessment for development of indicators?

There is a common misconception that the assessment, the analysis of assessment data and the resulting problem tree⁵⁹ only help with the development of the TOC and objective statements. On the contrary, the assessment results are tremendously useful for developing indicators. If you have challenges identifying indicators for SO(s), refer back to your assessment and problem analysis as they point to the main focus of the project, which in turn defines the change we want to see as result of the project. Assessment findings are also helpful for adjusting donor-mandated indicators. Refer to [Annex 5](#) for assessment-informed adjustment of a BHA mandatory indicator definition, placement in the Proframe, and selection of the most appropriate data collection method.

59. For more information on problem trees, consult [ProPack I](#) (CRS 2015a).

STEPS

STEP 1
REFLECTSTEP 2
RELATESTEP 3
PRIORITIZESTEP 4
REFINE**Number versus percentage: When to use what? Should we ever use both?**

A common rule of thumb is that indicators at the output level are worded as numbers while indicators at the IR or SO level are worded as percentages. But there is no requirement for this! Whichever you choose, be aware of the implications of selecting one or the other. An indicator worded as a number requires a census of the population, i.e., counting every single instance of the indicator being met. This is often only feasible at the output level when the indicator tracks the actual delivery of services or activities. An indicator worded as a percentage allows for sampling of the target population, which may be appropriate at any level of a results framework. For instance, the output-level indicator “*number of training participants who have essential knowledge about x*” requires that every training participant takes the knowledge test and reports on the actual number of participants who scored at a certain level. The same indicator worded as “*percentage of training participants who have essential knowledge about x*” allows you to test only a random sample of training participants. The choice depends on:

- ◆ What the indicator measures and how complex that change is. For example, measuring improved well-being is much more complex than measuring receipt of NFI kits. For the former, a percentage is more appropriate as it allows for sampling, while for the latter, an indicator expressed as a number is needed for compliance purposes.
- ◆ How large the focus population of the indicator would be and how feasible it would be to count or obtain this number for your denominator in a percentage calculation. For example, the indicator percentage of school-aged children enrolled in community-based education classes requires information on all school-aged children in all the locations we work in, which may not be feasible to obtain.
- ◆ Whether you may need to track instances from the whole population for other reasons. This most frequently happens for compliance purposes. For example, we need to report to the donor the number of households that received upgrades to their shelter. A percentage indicator and sampling of the population would not be appropriate here.

Note that some donors (e.g., BPRM) require reporting of both numbers and percentages for each indicator. Unless this is a donor requirement, it is strongly recommended that only one or the other is selected.

STEPS

STEP 1
REFLECTSTEP 2
RELATESTEP 3
PRIORITIZESTEP 4
REFINE**STEP 3: Prioritize the indicators with their use in mind**

Who	Programming and MEAL staff or a smaller group consisting of key programming and MEAL staff (e.g., project manager or chief of party, MEAL manager and, if available, senior partner staff).
Resources	List of indicators drafted in Steps 1 and 2, TOC and SOs, IRs and outputs.
How	<p>At the participatory session engaging either the larger group (as in the previous step) or the smaller group of key staff listed above, discuss the following:</p> <ul style="list-style-type: none"> ➤ Are all indicators developed through Steps 1 and 2 useful for program management? When and how will the team use⁶⁰ the information collected through the indicator? ➤ Reflect on how data for each indicator will be collected. Does any indicator require a level of effort or resources for data collection that the project cannot afford or that is not commensurate with the benefit it will bring the project (e.g., population-based surveys, multiple data collection points over time, etc.)? ➤ Review the indicators for an entire causal stream: SO-IR-Output. How similar or different are they? If they are very similar, can one be removed without losing important information? ➤ Will they generate enough information for adaptive project management? Will the indicators together enable testing of the TOC and generate learning about the if-then logic throughout the causal stream(s)?

Note that prioritization of indicators may continue beyond this initial session. It is recommended that the sector technical advisor (TA) of the sector is consulted to bring sectoral knowledge and experience to confirm whether the selected indicators are the most relevant and appropriate for the project. For example, a food security expert can help choose the most appropriate indicators for a food security project among the many options available in this sector.

60. Use may include reporting to donors and other project stakeholders, as well as project adaptive management and learning.

STEPS

STEP 1
REFLECTSTEP 2
RELATESTEP 3
PRIORITIZESTEP 4
REFINE

Frequently asked questions

How many indicators is enough?

Aim for a limited set of carefully selected indicators that respond to donor requirements and that together capture the core elements of the TOC and objective statements which they aim to measure. Ideally, you would have up to two indicators per each objective statement—SO, IR or output. While you can have as many indicators as you like attached to one statement, if you have three or more, the project MEAL system may become too cumbersome and overwhelming.⁶¹

What should we do if we have too many indicators associated with one objective statement?

Often the indicators measure sequential phases or specific activities in the achievement of an objective. Consider whether you can eliminate any process-oriented indicators or combine process-oriented features into a more comprehensive, outcome-focused indicator. For example, for the output *Social workers support families with immediate medical referrals*, the team identified two indicators: *Number of families identified with medical needs* and *Number of families receiving referrals for identified medical needs*. Project management needs will likely be better met by one indicator that combines the two project activities (and adds a timebound element to make it SMART): *Percentage of families identified as in need of medical support who have received a referral within 1 month*.

How do we decide among the many options for measuring one objective statement?

When you have many indicators measuring one objective statement, reflect on the following questions:

- ◆ Which indicator(s) will generate the most useful information for project management? The use of the data generated by the indicator should always be the driving factor in prioritizing indicators.
- ◆ Do all indicators focus on the core elements of the desired change, those that are most critical to the project's pathway of change? Do any indicators measure secondary or minor aspects of the desired change?
- ◆ Are any indicators “too easy”? Sometimes, indicators focus on aspects of the project that are almost completely within project control. This is often signaled by a 100% target. In this case, consider whether these can be removed, or whether they can be made more useful for project management, for instance by including a timebound element (e.g., “percentage of displaced households receiving winterization NFIs ...” “... before the onset of winter” or “... by December 15”).

If you are not able to decide which indicators to retain and which to remove at this stage, retain all the indicators and go through the next steps. You will have another chance to prioritize as you start refining and running the indicators through the SMART checklist.

61. CRS 2015a.

STEPS

STEP 1
REFLECTSTEP 2
RELATESTEP 3
PRIORITIZESTEP 4
REFINE**STEP 4: Refine the indicators and run them through the SMART checklist**

Who	Key or senior programming and MEAL staff, including relevant sector technical advisors.
Resources	List of prioritized indicators, close-to-final or final TOC, RF/Proframe, donor-defined PIRS for selected donor indicators.
How	Refine the final wording of the selected indicators. Be sure to include the structural elements of each quantitative indicator: unit of measure, subject of measure, disaggregation categories and description of change. Include a draft definition of broad or ambiguous terms mentioned in the indicator to capture project-specific context and nuances.

As the details of the indicators become clear, run your indicators through the SMART checklist ([Figure 1](#) above) to fine-tune them and ensure they have all the required dimensions. This step often happens in several rounds, as the TOC and RF/Proframe are finalized and as you start working through the details of your indicators using the design tools described in the next section.

Annex 5: Adjusting a donor-mandated standard indicator: The case of a BHA indicator

Donor-mandated indicators are an essential requirement in many proposals. While the level of adjustability depends on the donor and the indicator, in most cases, some flexibility exists in terms of indicator definition, its placement in the Proframe, and the method of data collection and analysis, etc. Some donors only list required indicators for a specific sector (e.g., BPRM), while others share a PIRS with all prescribed details (e.g., BHA). Navigating through donor guidelines and instructions and ensuring that donor-mandated indicators are useful for project management rather than being an added data burden is often a challenging process. Also, when provided, some PIRS details are either too methodologically poor or challenging to be put into practice.

This annex presents a step-by-step approach for adjusting a donor-mandated indicator, using the example of a BHA standard indicator:

Number of households occupying shelter that is provided pursuant to relevant guidance appearing in the Sphere Project Handbook

This indicator draws on the Sphere standards and has its own PIRS included in the donor guidelines. It is an example of the typical methodological challenges in donor PIRS that this guidance aims to address.

This step-by-step approach is applicable to all donor-mandated indicators. Note that most donors are much less prescriptive than BHA, and thus there is more flexibility in the way the indicator is defined and developed. The adjustment can be done at the time of proposal development, especially if the PIRS are required, or after the proposal is awarded and the Work Plan and refined MEAL plans are due.

Step 1: Review donor PIRS for mandated indicators and check the indicator definition

Donor PIRS specifies the following:

S12: Number of households occupying shelter that is provided pursuant to relevant guidance appearing in the Sphere Project Handbook.

Applicability	Required if applicable
Type	Output
Sector	Shelter & Settlements (S&S)
Sub-sector	Shelter

INDICATOR DESCRIPTION

Definition:

This indicator counts the number of households occupying a BHA-supported or -provided shelter consistent with Sphere Project guidance. “Shelter” refers to covered living space within a structure that provides:

- Adequate space and protection from cold, damp, sun, rain, wind or other threats to health;
- A location where essential household activities can be satisfactorily undertaken; and
- A location where livelihood support activities can be pursued, as required.

According to Sphere Project guidance, individuals should have sufficient covered living space to provide dignified accommodation, including, where possible and practical, minimally adequate covered living space of 3.5 square meters per person.

Step 2: Identify any terms that require further definition

In the definition above, there are a few terms that require further definition or clarification:

- “sufficient” covered living space
- “dignified accommodation”
- “minimally adequate” covered living space

Note that “sufficient” and “minimally adequate” covered living space has already been defined in the specification of 3.5 square meters per person in the PIRS and points 1-3 but further clarification is needed, as well as definition of “dignified accommodation”, as highlighted in the following steps.

Step 3: Identify donor-recommended sources of information to clarify the terms

Since this indicator draws heavily on Sphere, the definition of the above terms should be informed by the indicators related to the relevant Sphere standard. In this case it is:

Sphere Shelter and Settlement Standard 3, Living Space:

People have access to living spaces that are safe and adequate, enabling essential household and livelihoods activities to be undertaken with dignity.

Sphere key indicators for this standard are:

1. Percentage of the affected population who have adequate living space in and immediately around their shelters to carry out daily activities.
 - ◆ Minimum 3.5 square meters of living space per person, excluding cooking space, bathing area and sanitation facility.
 - ◆ 4.5–5.5 square meters of living space per person in cold climates or urban settings where internal cooking space and bathing and/or sanitation facilities are included.
 - ◆ Internal floor-to-ceiling height of at least 2 meters (2.6 meters in hot climates) at the highest point.
2. Percentage of shelters that meet agreed technical and performance standards and are culturally acceptable.
3. Percentage of people receiving shelter assistance that feel safe in their shelter.

BHA specifically mentions in the PIRS that Sphere compliance means ensuring a minimal covered living space of 3.5 square meters per person (wherever feasible). But we have the choice of whether to also include:

- **Technical and performance considerations** (disaster risk reduction, ventilation, insulation for hot/cold climate, environmentally sustainable materials)
- **Culturally appropriate** (this often includes gender considerations related to privacy, but also shelter siting, design and layout)
- **Personal and physical safety** (e.g., doors and windows, locks, privacy/protection, accessibility to people with disabilities or the elderly, etc.)

Step 4: Prioritize the aspects of the definition that your project focuses on

To do so, we need to consult the **assessment data**. The definition should be refined to focus on key priority needs or gaps identified during the assessment. Let's examine two different scenarios to illustrate how the assessment affects the indicator definition and its adjustment.

Scenario 1

The assessment showed that the majority of households affected by the crisis are living in overcrowded spaces with several families together. This was the one overwhelming concern mentioned by all respondents across disaggregation categories. Thus, our indicator definition can focus exclusively on space per person.

The adjusted definition would be:

Definition: This indicator counts the number of households occupying a BHA-supported or -provided shelter consistent with [Sphere Project](#) guidance. Based on assessment results [reference the assessment report if needed], this indicator will focus on ensuring covered living space within a structure that provides adequate space to all family members, defined as 3.5 square meters per person.

The indicator would likely fit in at the SO level since it measures physical improvement of shelter as well occupancy which is an end-of-project achievement.

Scenario 2

The assessment also reported protection concerns, specifically that women complain about the lack of private space in the shelters in which they can be secluded from other family members and from visitors. In this case, the indicator definition should also include privacy and personal safety considerations.

The indicator definition would be adjusted in the following way:

Definition: This indicator counts the number of households occupying a BHA-supported or -provided shelter consistent with [Sphere Project](#) guidance. Based on assessment results [reference the assessment report if needed], this indicator will focus on:

1. Adequate covered living space, defined as 3.5 square meters per person, and
2. A location that offers women adequate privacy and protection, self-assessed by female members of household (using a 5-point Likert scale ranging from “not at all safe” to “very safe” in their shelter; responses “safe” and “very safe” will be counted in the indicator).

For the purposes of adaptive project management, each dimension of this indicator will be tracked separately, but will be reported as one combined indicator.

As above, the indicator would likely be at the SO level in the results framework, as it also represents a qualitative improvement in the shelter situation and the end-of-project change our project aims to deliver.

Potential challenges: The second dimension of the indicator includes a scale in its definition, while it indicates the need for individual interviews with female household members. How feasible is this for an indicator that is a number? What are some ways in which this challenge can be addressed? Please see below.

Step 5: Adjust other details of the indicator

Once the definition is clarified, go through the rest of the PIRS and see whether anything else needs to be adjusted. The rest of the PIRS says:

Unit of measure: Number (of households)

Calculation: This is a count of households occupying the shelter that is provided by the activity pursuant to relevant guidance appearing in the Sphere Project Handbook.

How to count Life of Award (LOA): LOA values are the reported values at the end of the award, counting only the unique number of households, without double counting, who are occupying shelter that is provided by the activity pursuant to relevant guidance appearing in the Sphere Project Handbook.

Direction of Change: +

Disaggregation: Gendered household type: Female and male, female no male, male no female, child no adult

DATA COLLECTION

Method: Routine monitoring

Source: Monitoring checklist/form

Who collects: Implementing partner staff

From whom: Beneficiaries

Frequency of collection: Data will be on an ongoing/rolling/monthly basis.

Frequency of reporting: Data will be reported in the semi-annual report, annual report and final performance report.

Baseline value information: Baseline value is zero.

Note that “routine monitoring” is technically not a method, nor does “monitoring checklist or form” represent a data source (this is a tool). This technical ambiguity represents an opportunity to clarify and contextualize the method and source, so it suits your project situation. Let’s return to the two scenarios to see how the teams addressed these two details in the PIRS that was resubmitted to the donor after the project was awarded.

Scenario 1

To measure the number of shelters that provide adequate space of 3.5 square meters per family member, we may propose the following methods for data collection:

Record review (least level of effort): If shelter is provided in kind per standard design or through contractors

- ◆ Shelter distribution data on the number households who received shelter with BHA funding.
- ◆ Technical design / bill of quantity to confirm shelter size per family (**).
- ◆ Registration data on household size (to determine whether condition of 3.5 square meters per household member is met).

*** Depending on shelter delivery modality (e.g., if self-built), you may need to verify actual shelter size. This could be done either through self-reporting as part of a household survey or through direct observation of a sample of shelters.

You may want to triangulate record data and verify actual occupancy of the shelters. For instance, you could visit a sample of shelters to confirm distribution records data and, through **direct observation**, verify that all shelters are indeed occupied. You could also validate the reliability of your registration records on family size through **self-reporting** through **structured interviews/surveys** on the actual number of people living in a random sample of shelters. Triangulation of record data through direct observation and household interviews on a small random sample of target households would be sufficient to validate the use of distribution and registration records for reporting.

Scenario 2

This indicator is more complex as it comprises two dimensions (occupancy and privacy/safety), each using a different data source. To measure the second dimension of this indicator, one more monitoring method would need to be added: **interviews** of a random sample of female family members about their sense of safety in the shelter. The data collection tool will be a questionnaire, with specific questions on privacy and safety, where female family members would assess it using 1-5 Likert scale. As noted above, this indicator is a number, which therefore requires a census, i.e., data from an entire population.

How challenges were addressed: In both scenarios, the team proposed a verification process:

- Scenario 1: Triangulation of record data through direct observation and household interviews with a small random sample of target households to validate the use of distribution and registration records.
- Scenario 2: A structured interview with a sample of female household members to verify whether the shelter quality conditions (Sphere compliance) have been met.

In both cases, the rationale behind this approach is that if sampling is appropriate (i.e., random sampling of households), it should be representative of the larger population of the area, hence conclusions on occupancy, privacy and safety can be generalized to the broader population.

Main tips to remember

- Assessments are key to informing indicator definitions and contextualizing donor-mandated indicators. Decide which aspects of the definition to emphasize.
- Founding indicators on assessments ensures their relevance and use.
- Think about possible data sources and methods, and implications for data collection (timing/frequency and level of effort). Each decision on adjustment of the standard indicator has implications on data collection details.
- For multi-dimensional indicators, track progress against each component separately, for problem-solving purposes. Consolidate the data when reporting to the donor (in the IPTT) but discuss in the narrative any significant differences between components.
- You can adjust donor-mandated indicators even when the PIRS is prescriptive. If you can justify the adjustments, the donor will often accept them.
- When deciding on and adjusting industry standard indicators, always consult your sector TA.

Annex 6: PIRS template⁶²

[PROJECT]

Objective: [state the objective]**Indicator:** [state the indicator]**INDICATOR DESCRIPTION****Indicator definition** [please define all ambiguous terms in the indicator]:

Remember:

- All ambiguous terms in the indicator need to be defined. Definitions help to clarify the indicator and show it is relevant to our project. They provide a good opportunity to make donor indicators work for you: to operationalize and contextualize them.
- The definition should be detailed enough to ensure that if different people at different times are given the task of collecting data for a certain indicator, they will collect identical data.⁶³
- The definition should be simple, explaining or specifying each complex term or dimension within the indicator separately, but also explaining how all dimensions work together. Make sure that the definition avoids ambiguous terms, e.g., in the indicator *number of students engaged*, “engaged” cannot be defined as “actively involved.” Try to find more concrete terms or criteria that will be understood by all project and partner staff in a same way, for instance “engaged” may be defined as being present at a minimum of 3 out of 4 meetings per month.
- Indicator definitions should be developed by programming staff, often with input from technical/sectoral experts. Definitions are often informed by the assessments (Sections 5 and 7a).
- How the indicator is worded may have implications for data collection, e.g., how many data points are needed to calculate an indicator (Section 7a and percentage calculation conundrum).

Numerator and denominator: [in the case of a percentage or ratio, specify the formula for calculation]

Remember:

- Beware of how the indicator is worded. For example, *percentage increase in knowledge* does not have the same formula for calculation as *percentage of students with increased knowledge*.
- The phrasing may have implications for your data collection (e.g., how many data points are needed to calculate an indicator) and sampling (e.g., whether you need to sample the same respondents or not).

Unit of measure: [specify unit of measure for the indicator]

This is typically households or individuals, but it can also be schools, health centers, committees, villages or municipalities, etc. Every quantitative indicator has a unit of measure.

62. The template presented here is adapted from [Bureau for Humanitarian Assistance](#). Check donor requirements before completing the PIRS.

63. USAID 2010b.

Disaggregated by: [specify key disaggregation categories]

Remember: Disaggregated data help track whether or not specific groups participate in and benefit from project activities.⁶⁴ Disaggregation categories inform the analysis, so it is vitally important to choose relevant subgroups that are meaningful and will help you better understand project progress and support adaptive project management. At a minimum, all indicators should be disaggregated by male and female respondents (when applicable).

Suggested data collection method [specify data collection method; and tool, if known]

Remember:

- The method and tool need to match the indicator type. Quantitative indicators call for quantitative methods and tools, qualitative indicators require qualitative methods and tools.
- Remember that *method* and *tool* are not the same. The method is *how* you plan to collect information, e.g., individual structured interview, focus group discussion, direct observation, while the tool is the *form* you will be using to collect that information, e.g., survey questionnaire, FGD guide, observation checklist.

Suggested data source and sampling: [specify data source]

Remember:

- The data source may be people (e.g., project participants, children aged 13-18, women) or records (e.g., existing project records, copy of government decision). Data sources are not always the same as the unit of measure, for instance, caregivers may be the data source for children's well-being.
- Sampling for quantitative indicators needs to use random sampling techniques (e.g., simple random sampling, stratified, cluster, etc.), while qualitative data requires non-random or purposeful sampling (e.g., best-worst case, typical case, snowball, etc.).⁶⁵

Timing and frequency of data collection: [specify timing, duration (start/end) and frequency]

Remember:

- The start and end, as well as the frequency of data collection needs to be appropriate to the project dynamics (as specified in the detailed implementation plan (DIP)) and the time expected for a certain change to occur. Often data for indicators measuring lower levels of the Proframe (outputs) starts earlier and is collected more frequently than for indicators at higher levels (IR and SO). Change depicted at the IR and SO levels and measured through IR and SO indicators is usually more complex and takes time.
- If your project is seasonally sensitive, be sure to take that into consideration as you think through the frequency of data collection, otherwise, your data may be incomparable.
- In most cases, only data collection frequency is mentioned in the submission to the donor, but it is strongly recommended that you also document timing (when in the calendar year) and start-end in internal PIRS.

64. USAID 2010b.

65. For more guidance on sampling, refer to [Guidance on Monitoring and Evaluation](#) (CRS 2012)

Analysis and use of data [specify timing, responsibility, analysis approach and use]

Remember:

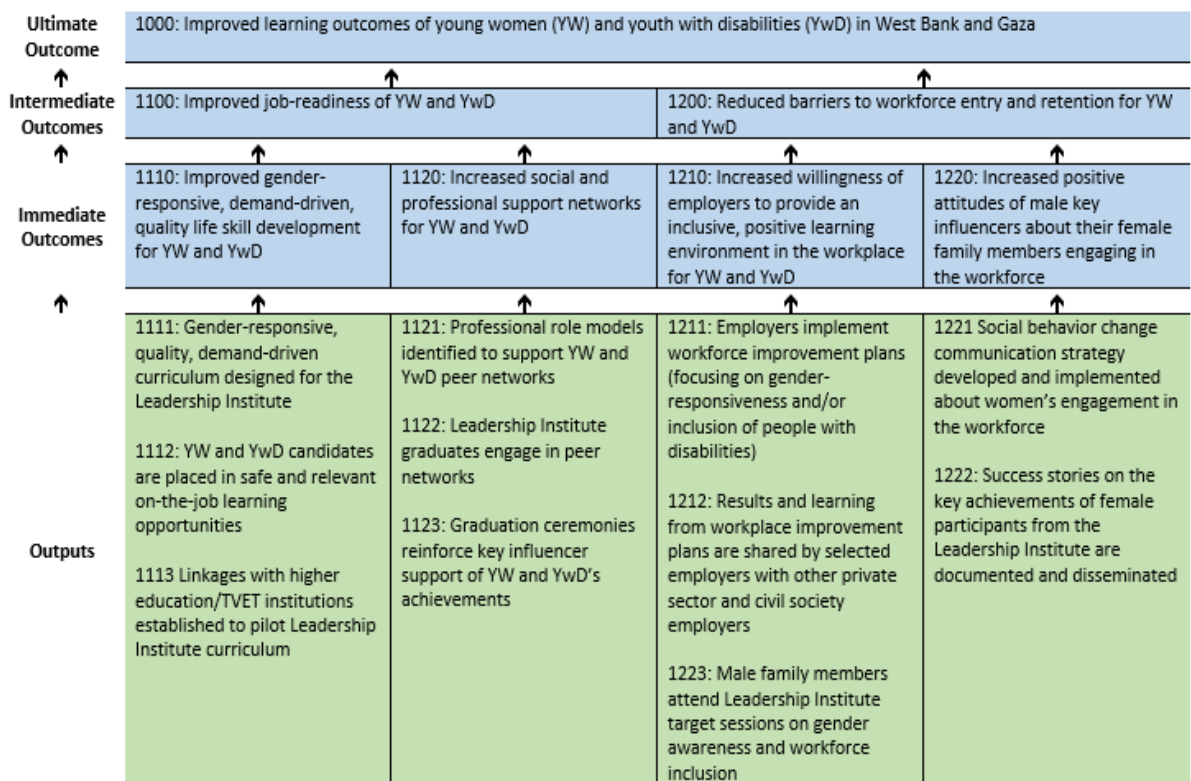
- The frequency of data analysis does not always match the frequency of data collection. In some cases, you may collect data on a monthly or even daily basis, but conduct analysis only on a monthly or quarterly basis. Remember, you cannot analyze data more frequently than you collect it.
- Be as specific as possible when describing who is responsible for the data analysis.
- The method of analysis can be fairly general at the proposal stage. It is based on single-indicator plans for which the PIRS is developed and uses terms such as “thematic analysis” for qualitative data, or “descriptive statistics,” “comparison against baseline and target,” or “inferential statistics” for quantitative analysis. More detailed planning for analysis should take place during the SMILER+66 when the MEAL operating manual is developed, including a learning plan.
- Remember: the use of data is not limited to reporting. Data should be used to validate assumptions, adapt activities and project strategies, check the TOC and respond to learning questions, thus enhancing project-based learning.
- Include relevant project review meetings (e.g., quarterly, annual, midterm, etc.) when data will be used to make programming decisions.

Annex 7: Salalem complex indicator and how we learned to love it

Introduction

Salalem is a five-year gender-transformative and disability-inclusion project targeting over 7,000 participants, including young women, youth with disabilities and key male influencers within their households in Gaza and the West Bank. The project goal focuses on “improving the learning outcomes and employability of young women and youth with disabilities in West Bank and Gaza.” At the heart of the approach is the Leadership Institute, a three-month livelihoods training with foundational life skills, followed by either job readiness or entrepreneurship training. The project is grounded in the Social Ecological Model (SEM), which suggests that a transformative approach needs to engage multiple levels of society to produce change. As such, Salalem is expected to indirectly engage over 20,000 community members through complementary project components that integrate families, employers and the community at large.

Figure 1. Salalem logical framework



Within Salalem, there is a heavy emphasis on piloting and learning. By piloting components of the project before bringing them to scale, the team has been able to identify what is working and what needs to be refined further, both in the implementation of project activities and in the way project success is measured. This case study outlines Salalem’s lessons learned about measuring one of the project’s three Ultimate Outcome indicators:⁶⁷

67. “Ultimate Outcome” is a Global Affairs Canada (GAC) term for “goal.” Unlike many donors, GAC requires indicators for the project goal.

The percentage of female leadership institute graduates (including young women with disabilities) who report a high or very high sense of agency (4 or 5 on a 5-point scale).

- The performance monitoring plan (PMP) noted: “Agency” to be defined through participatory activity at project start. The definition may include but is not limited to: “ability to make own decisions related to workforce engagement,” and “more equitable voice in household decisions.”

How is “agency” measured?

As the team was embarking on defining this indicator, they encountered several key challenges:

- **Agency is intangible.** Agency cannot be weighed, counted or even seen. How does one define “sense of agency” so it is specific and supports objective measurement that can be replicated over time, especially knowing that a sense of agency is grounded in self-perception?
- **Agency is a multi-dimensional construct.** Agency is complex because it is multi-dimensional. We exercise agency in different ways, in different parts of our lives and at different moments of our lives. How do we ensure our definition and our data collection tools capture those dimensions, and help respondents recall situations and moments of their lives when they exercised agency (or not)? How do we ensure the definition focuses on the aspects of agency relevant to the project focus?
- **Agency is a new or emerging concept in the target communities.** At the start of the project, it became clear that the concept of agency—for both men and women—was novel. How can we introduce a new, multi-dimensional concept and ask respondents to self-report on it?

These challenges are ubiquitous in sectors that aim at changing social or cultural norms. Such projects have indicators that measure intangible, multi-dimensional/composite concepts that are either new or emerging, and often less acceptable in targeted communities.

Community consultations

As noted in the approved performance monitoring plan, the team planned to solicit community input into the definition of “agency.” Through this process, the team had hoped to **contextualize** their understanding of what “agency” means to key stakeholders in Gaza and the West Bank. Since during the negotiations with the donor, the indicator definition became pretty prescribed, the community consultations were aimed at informing rather than fully developing this indicator.

Integrate MEAL into project activities

Remember that there are ways to integrate MEAL into project activities or processes that have already been planned or are in place; this helps create efficiencies for our teams and the respondents who are graciously giving us their time.

The team looked at the detailed implementation plan (DIP) to identify efficiencies in early start-up, as well as to identify other information needs and gaps to capitalize on this data collection process. As a result, the team came up with two key objectives for community consultations:

- Identify skills gaps to inform the curriculum design decisions.
- Help to inform and contextualize key indicator definitions such as “agency.”

The team designed data collection tools—focus group discussion (FGD) guides—to meet these information needs. Through FGDs with over 80 young women, youth with disabilities and male influencers the team learned the following about the community definition of “agency”:

- The respondents validated that agency is closely linked to decision-making.
- The respondents noted the important point that agency does not mean making decisions alone.
- The respondents highlighted the importance of freedom from discrimination.

Baseline data collection

To create efficiencies between the project implementation and MEAL activities identified through the DIP review, the team decided to integrate the **baseline data collection into the application process**. The application process was one of the first field activities in which all community members that met the project criteria were invited to apply for participation in the project by submitting their basic information and undergoing a short interview. Because the pilot phase of the leadership institute targeted “positive deviants” (those most likely to succeed), integrating baseline data collection into the application process would not only be efficient, but would also help get a more representative picture of respondents beyond the positive deviants that would be ultimately selected for participation in the project. The baseline tool included the following question to generate data for this indicator: *To what extent do you agree or disagree with the following statement: I feel like I have the agency to make decisions for my life.*

The question was strategically placed toward the end of the interview, after rapport was built through questions that explored the applicants’ demographics, work experience and future plans. To be responsive to the community feedback on agency, including co-decision-making, the response options included the explanation as specified in the table below. During the analysis, these were transformed into a level of agency ranging from low to very high as presented below:

Multiple-choice options	Level of agency
Strongly agree, I am the primary decision-maker in my own life.	Very high
Agree, I make most decisions about my life in consultation with my co-decision-maker.	High
Somewhat agree, I make some decisions about my own life, but most decisions are made by my co-decision-maker.	Moderate
Disagree, my co-decision-maker makes decisions for me.	Low
I decline to answer.	N/A

Results: We hit our target before we even started

The draft baseline report reads:

“During data collection from the pilot cohort applicants in Gaza, 96% (158 out of 164 applicants who were eligible for the project) reported a **high or very high sense of agency (“agree” or “strongly agree” response)**. The question and results for this indicator are detailed below.

“To what extent do you agree or disagree with the following statement: I feel like I have the agency to make decisions for my life.”

Strongly agree, I am the primary decision-maker in my own life.	74
Agree, I make most decisions about my life in consultation with my co-decision-maker.	84
Somewhat agree, I make some decisions about my own life, but most decisions are made by my co-decision-maker.	6
Disagree, my co-decision-maker makes decisions for me.	0
I decline to answer.	0

How can this be?

A closer look at and interpretation of the baseline results pointed to few possible issues in the data collection process:

Bias

In general, self-reporting is susceptible to under- or over-reporting depending on what respondents believe CRS is looking for in terms of criteria to participate in or to gain additional benefits. By asking this question within the application, which was not anonymous, the respondents may have been influenced by what they thought might make them more likely to be selected.

Unbalanced and insufficiently nuanced scale

The answer options were heavily weighted toward positive responses, with three options representing nuances of “agree” and only one option categorized as “disagree.” Also, the scale was too small to capture potentially nuanced perceptions that are often vital for concepts that are new or different from social norms. Finally, by adding “decline to respond” the team tried to acknowledge and recognize the sensitivity of the question, but, in so doing, one whole rating option was lost on an already a small scale.

The Salalem project manager reflected that “MEAL is both an art and a science. There were opportunities for us to have strengthened the design of this data collection but there were also decisions we made for which we had a good justification. This is part of the learning, piloting and testing process that is so important.”

Simplification of a complex concept

Despite knowing that agency is a complex concept, the tool asked only one question to gauge the sense of agency. On the one hand, this simplified the calculation of the indicator but, on the other, it resulted in having only one data point and no way to triangulate or create nuance in the analysis and interpretation of the responses.

Follow-up: Second round of baseline data collection

The team considered changing the indicator calculation to include only the top rating of “strongly agree.” This would have significantly decreased the baseline (to 46%); but it would not have addressed the other shortcomings identified with the data collection tool. The team decided to repeat the baseline data collection with a revised tool that would aim to capture the richness and nuances that they were looking for. The team made the following changes:

Literature review

With technical support from Gender and MEAL advisors, a literature review complemented the community consultations and clarified further potential dimensions of “agency,” i.e., goals, control and influence. The literature review also introduced the concepts of independence versus autonomy.

The new stand-alone baseline survey:

- Was anonymized to mitigate bias.
- Included an expanded 10-point response scale with response options ranging from “strongly agree” to “strongly disagree,” to allow more nuance in the responses.
- Ensured the scale was balanced between positive and negative response options.
- Included supplementary questions to deepen the understanding of “agency” and enable triangulation of responses.

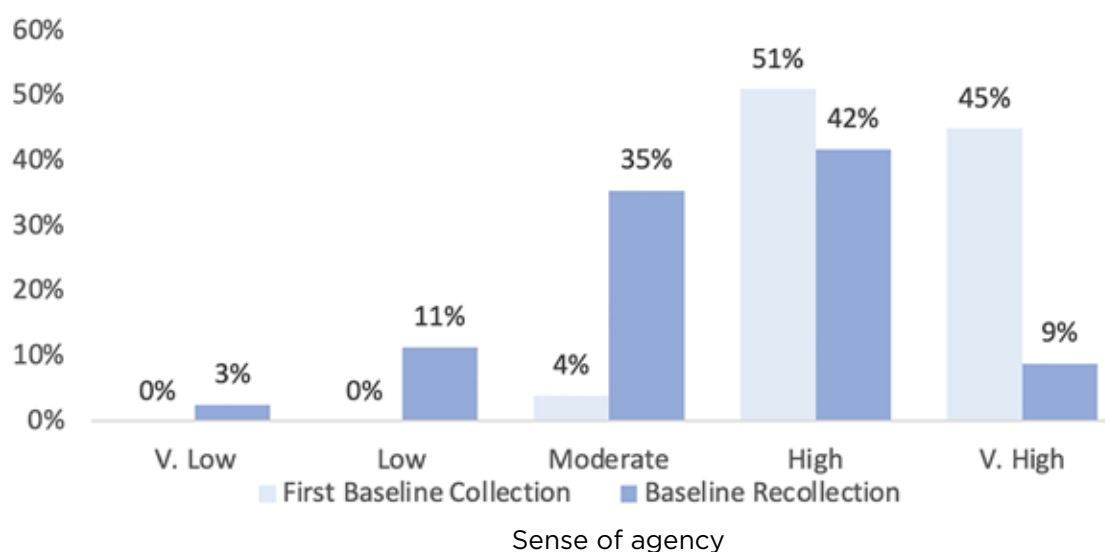
As in the first baseline data collection tool, the question on agency was listed at the end of the survey, but the questions that preceded it explored the different dimensions of agency in the respondents’ lives, thus helping them internalize the concept and frame their opinion for the final question that consolidated the concept.

Assumptions

The revised survey also included a question on discrimination (against women and people with disabilities), as this had been suggested in community consultations. While this was outside of the project scope or control, it was an important factor that could have influenced project success and therefore needed to be monitored.

The results of this repeat baseline were significantly different from the first one, and much more nuanced. Most significantly, they were much more informative for evidence-based project management.

Figure 2: Results of first versus second round of baseline data collection



Lessons learned

Baseline data collection is not only about meeting donor requirements

The baseline helped test the team's assumptions. For example, one of the common assumptions in the context of the West Bank and Gaza is that men—fathers, husbands and brothers—are the key influencers in the household. However, while over 50% of respondents reported that their fathers indeed had a high degree of influence in their decision-making, 65% reported that their mothers also had a high degree of influence (note that the categories were not mutually exclusive). This challenged the initial assumption and pointed to a different target population within the project to ensure key influencers in the household supported women to engage in the workforce. Based on the findings of the second baseline collection, the Salalem team developed a learning plan focusing on further exploration of household dynamics to test the initial assumption about the dominance of male influencers. This learning will further inform targeting of influencers within the households and guide the project's communication strategy.

The literature review and community consultation on the indicator definition go hand-in-hand

This is especially the case in projects that aim to transform prevailing cultural and social norms. The communities may not be aware of various aspects of the expected change, hence will not bring it up during consultations.

Community consultations help sharpen project focus and identify other factors that could impact success

Community consultations revealed an environment free of discrimination as a key factor for women when they considered their sense of agency. While the project scope was limited in its ability to impact discrimination on a large scale, the team identified it as an important project assumption that needed to be monitored throughout the life of the project. Also, this finding further enabled the refining of the community-level social behavior change activities and informed the content on which the training would focus, i.e., Palestinian Labor Law and “knowing your rights.”

Sensitive indicators require careful survey design

For indicators that measure sensitive issues, every detail counts, including how the questions are asked, what scale is used and where in the data collection tool they are placed. When exploring complex, abstract concepts, it is helpful to start with questions exploring more concrete sub-dimensions and then end with the final question exploring the actual concept. This approach allows the respondent to personalize and internalize the concept, reflect on its various components, and then frame their final opinion. The scale should be balanced and nuanced enough to pick up potentially small differences in opinions.

Keep it simple yet informative, reliable and useful for adaptive project management

While only one question directly asking about “sense of agency” was used for calculation of the indicator, the supplementary questions sought additional data critical for the triangulation of responses.

Such an approach helped avoid a complex scoring process, and provided sufficient data for verification and triangulation. Also, the team conducted a detailed analysis of the responses and scores on each specific sub-dimension, i.e., goals, control and influence. The scores per sub-dimension will be critical for adaptive management purposes as each sub-dimension has a specific set of activities associated with it.

Figure 3: Question on “sense of agency”

Please tick (✓) a rung on a ladder to represent your overall sense of agency:

Instructions: the bottom rung “1” is “I feel I have no control over my decisions and actions” and the top rung “10” is “I have full control over my decisions and actions.”

10	<input type="checkbox"/>	Very High
9	<input type="checkbox"/>	High
8	<input type="checkbox"/>	Moderate
7	<input type="checkbox"/>	Low
6	<input type="checkbox"/>	Very Low
5	<input type="checkbox"/>	None
4	<input type="checkbox"/>	
3	<input type="checkbox"/>	
2	<input type="checkbox"/>	
1	<input type="checkbox"/>	

And finally, be creative!

The team developed a “ladder” to visualize the scale (Salalem means “scale” in Arabic) (see Figure 3).

References

- Bohl** N, Silajdzic V and Ryan M. (2018). [MEAL4Kids: Guidance on measuring children’s well-being in emergency response and recovery programming](#). CRS.
- BPRM**. (2021). [General NGO guidelines](#), Appendix C: Programmatic Sectors, Modalities, and Standardized Indicators.
- Bradburn** NM, Sudman S and Wansink B. (2004). *Asking questions: The definitive guide to questionnaire design—For market research, political polls, and social and health questionnaires, Revised Edition*. San Francisco: JosseyBass Wiley Imprint.
- Church** C and Rogers MM. (2006). [Designing for results: Integrating monitoring and evaluation in conflict transformation programs](#). Search for Common Ground.
- CIVICUS**. [Monitoring toolkit](#).
- Cobo** A, Dolphin H and Prieto S. (2019). [Guidelines for de-identifying data](#). CRS.
- CRS**. (2010). [GAIN peacebuilding indicators](#).
- CRS**. (2015a). [ProPack I: The CRS Project Package; Guidance on project design for CRS project and program managers](#).
- CRS**. (2015b). [Community-defined Indicators guidance sheet](#).
- CRS**. (2019). [Social cohesion indicators bank](#).
- CRS**. (2020). [SMILER+ guide to MEAL system development](#).
- CRS and Humentum**. (2019). [A guide to the MEAL DPro](#).
- Cypress** B. (2017). [Rigor or reliability and validity in qualitative research: Perspectives, strategies, reconceptualization and recommendations](#). Dimensions of Critical Care Nursing 7/8, Vol. 36, Issue 4, p 253-263.
- DRL**. (2020). [Proposal submission instructions for applicants](#).
- Dzino-Silajdzic** V. (2018). [Practical guide to focus group discussions](#). CRS.
- Dzino-Silajdzic** V. (2020). [Practical guidance on developing a project’s theory of change](#). CRS.
- Global** Affairs Canada. (2016a). [Results-based management for international assistance programming at Global Affairs Canada: A how-to guide 2nd Edition](#).
- Global** Affairs Canada. (2016b). [Feminist International Assistance Policy Indicators](#).
- Hagens** C, Morel D, Causton A and Way C. (2012). [Guidance on monitoring and evaluation](#). CRS.
- Hind** J. (2010) [Additionality: A useful way to construct the counterfactual qualitatively?](#) *Evaluation Journal of Australasia*, Vol. 10, No. 1.
- Maxwell** D, Coates J and Vaitla B. (2013). [How do different indicators of household food security compare? Empirical evidence from Tigray](#).
- Morel** D, Silajdzic V and Hagens C. (2019). [Practical guide: Monitoring for problem-solving, adaptive management, reporting and learning](#). CRS.
- Parsons** J, Gokey C and Thornton M. (2013). [Indicators of inputs, activities, outputs, outcomes and impacts in security and justice programming](#). Department for International Development and UK Aid.
- People in Need**. (2021). [IndiKit: Rapid guide to designing SMART indicators](#).

- Sphere.** [Sphere Humanitarian Standards.](#)
- SurveyMonkey.** [Learn when and how to use Likert scale survey questions.](#)
- USAID BHA.** (2020). [Emergency Application Guidelines, Annex B: Indicator list.](#)
- USAID BHA.** (2020). [Indicator handbook for emergency activities.](#)
- USAID BHA.** (2020). [USAID-BHA ITT suggested template.](#)
- USAID.** (2010a). [TIPS: Baselines and targets.](#)
- USAID.** (2010b). [TIPS: Preparing a Performance Management Plan.](#)
- USAID.** (2010c). [TIPS: Selecting performance indicators.](#)
- USAID.** (2020a). [Food for Peace indicators handbook, Part I: Indicators for baseline and endline surveys for Development Food Security Activities.](#)
- USAID.** (2020b). [Food for Peace indicators handbook, Part III: Indicators for emergency activities.](#)
- USAID.** (2021a). [BHA indicator handbook, Part II: Monitoring indicators for Resilience Food Security Activities.](#)
- USAID.** (2021b). Program Cycle. [ADS 201 Additional Help: Disaggregating Monitoring Indicators.](#)
- USAID.** (2021c). [USAID Bureau for Humanitarian Assistance technical guidance for monitoring, evaluation and reporting for Resilience Food Security Activities V2.0.](#)
- USAID.** [Standard Indicators.](#)
- WHO.** (2015). [Global Reference List of 100 Core Health Indicators.](#)



Catholic Relief Services | 228 W. Lexington Street, Baltimore, MD 21201, USA |
crs.org | crsespanol.org