Guidance on Results Frameworks, Performance Indicators and Monitoring for Sustainability

**Based on a synthesis review of
OTLA project evaluations**

Josh Meuth Alldredge

Sarah Liuzzi

September 18, 2020

A. Guidance on theories of change and results frameworks

1. Challenges in developing theories of change and results frameworks

OTLA requires its grantees to express project theories of change (TOCs) in results frameworks (RFs). In most projects reviewed for this synthesis, ILAB developed rough TOCs before the project was awarded and required implementers to use the TOC. Midterm evaluations often recommended updated TOCs, and implementers developed new RFs while project operations were ongoing. However, USDOL's Grant Officers do not allow project objectives and high-level outcomes to be changed after grants are awarded, so changes to the corresponding inputs and outputs have led to disjointed RFs (because contexts evolve and priorities may shift, original RFs tend to define project objectives broadly).[1] As a result of this challenge and other factors described in the main synthesis report, evaluators (and reviewers for this synthesis) found the quality of some RFs and TOCs to be low, as shown in Figure C.1. The most common deficiency in TOCs was that they were not an accurate reflection of the projects they were meant to describe, reflecting unreasonably ambitious goals given the resources and time available for project inputs. The most common deficiency in RFs was having an unclear causal chain from input to output, and output to outcomes.

Figure C.1. Reviewer assessment of TOCs and evaluators' assessments of RFs

TOCs reviewed in synthesis (of 19)	
Low-quality TOC	2
Medium-quality TOC	13
High-quality TOC	4

RFs examined by evaluators (of 19)	
Inadequate RF	9
No mention of RF	6
RF mentioned but not assessed	3
Adequate RF	1

These challenges in TOCs and RFs have implications for project success. Our review of the 19 project evaluations suggests that projects with the strongest evidence-based, theory-driven approaches[2] achieved moderate to high effectiveness in reaching planned goals, whereas projects with poorly formulated theories of change had mixed effectiveness. Donors may wish to strengthen their support to implementers in their work developing evidence-based theories of change and express them in coherent results frameworks.

2. Best practices for developing strong results frameworks and theories of change

An RF visually structures a TOC to display the chain of assumed causal effects. Given that the RF provides orientation for the entire project effort and allows for the measurement of performance, donors and implementers should adhere to best practices, including the following:

---

[1] ILAB has indicated that it draws broader scopes at project conceptions, so that there is room to later shift and incorporate work that is deemed necessary to achieve the project objective. If scope is too narrow, leaves little to no room to elaborate results as implementation proceeds based on evolving context, shifting priorities, needs and understanding of capacities and power dynamics. However, these broad scopes can also generate the unreasonably ambitious goals that evaluators cited as a common problem in theories of change.

[2] To assess the evidence base and theoretical backing of the TOC for each project, we reviewed evaluators' interpretations of projects' use of evidence and theory in developing their interventions. We scored the degree to which the project successfully deployed evidence and theory on a scale of 0 (poor) to 2 (good).

- Changes in each level of the RF (e.g. inputs/activities) contribute to changes in the level above (e.g. outputs) through **strong causal linkages**. These causal linkages **must be explicit and, when possible, justified using concrete evidence from prior studies or programs.** Evidence from prior studies or programs may be documented in a narrative to accompany the RF that identifies the evidence for linkages as well as linkages for which the grantee has not found evidence. We recommend that RFs begin at the activities and inputs level to capture the full causal chain from the implementation of activities and inputs to the long-term outcomes.

- Each linkage depends on internal and external **assumptions, which should be made explicit, interrogated, and assessed** for the level of threat they pose to the project if the assumption is false. Interrogation of assumptions is a key step in identifying project risks. Higher risks are present when assumptions are less likely to hold; assumptions that are more likely to hold imply lower risk. The process of identifying and interrogating assumptions allows USDOL and implementers to anticipate risks and proactively develop mitigation strategies.

- Results at the same level (such as all outputs) must be **"individually necessary and jointly sufficient to achieve the level above them"** (such as outcomes) (USAID 2018). This means at each level, donors and implementers designing RFs should carefully examine every item for its value and necessity to the causal chain, and then carefully examine each level to ensure the items within it are adequate to drive expected change at the next level.

In Figure C.2, we provide an example of a single causal chain from an RF. It demonstrates examples of the best practices discussed above.

Figure C.3. Example causal chain from an RF
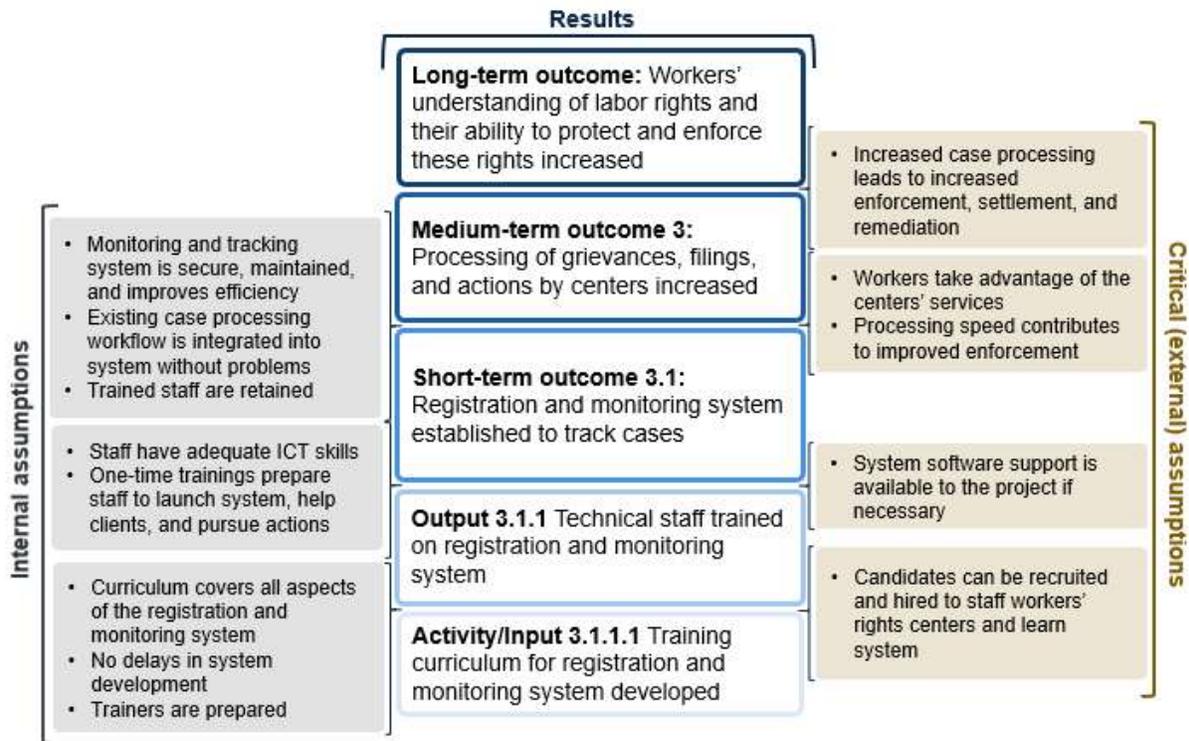
Sources: Interim Performance Evaluation of WRC, IMPAQ, 2019; and Technical Note: Developing Results Frameworks, USAID, 2018

This sample RF shows the causal linkages between each level, as well as example assumptions that might underpin the linkage.

As noted in the body of this report, projects focused on building government capacity were less able to achieve their goals than projects targeting workers or employers. One of the core reasons for this difference is that donors and implementers appeared to assume that project components would be sufficient to achieve substantial improvement in outcome areas and produce lasting change at the development objective level. However, these assumptions often fell through. To make the RF a more reliable guide, projects should interrogate assumptions early in the results framework development and whenever the framework is revised. **This implies conducting a needs assessment and stakeholder analysis at the project design phase.**

Interrogating assumptions and identifying risks means asking questions like:

- "What might happen to the buy-in of the MOL if a new political appointee arrives?"
- "Will employers really be enthusiastic about this training taking place with their workers?"
- "How exactly do we anticipate our trainings will change the behaviors of workers? What might get in the way of that change?"
- "What incentives might get in the way of government officials doing better inspections?"

Explicitly listing the expected causal links and interrogating the risks and assumptions in an RF can help implementers foresee challenges. Implementers must consider and list all assumptions required at each causal link. This step should be followed by carefully planning activities to 1) address holes discovered in the RF and 2) preemptively mitigate challenges and plan for risks that could materialize if the assumptions they have identified do not hold. Development of the RF is a tool to facilitate this process.

Donors may also exert more effort in developing a strong evidence-based theory of change and codifying it in a results framework during project concept and solicitation, even before the implementer is selected. When feasible, donors may wish to populate the RF with a set of indicators that include standard ones used across projects to facilitate cross-project comparison. In conceptualizing a project before solicitation, donors should review relevant academic and gray literature to assess the strength of evidence for the linkages described in the proposed project's RF. If the donor has already assessed the evidence base, the implementer may then review the donor's work rather than beginning it "from scratch."

For projects that work with government stakeholders, the assessment of risks and interrogation of assumptions should include elements of **political economy analysis**, assessing stakeholders' levels of interest, incentives, and power and the relevance of each for the implementation of the project.

B. Guidance on indicators

Some PMPs reviewed for this synthesis had poorly developed indicators and heavily revised targets, limiting the indicators' usefulness for analysis. In this section we present the characteristics of strong indicators, a sample of poor indicators and improved versions of them, and a selection of key indicators that may be relevant for most implementers.

Indicators help implementers and donors track progress toward desired inputs, outputs, outcomes, and ultimate outcomes. Targets set for indicators should be specific, measurable, achievable, relevant, and time-bound (SMART), and the target-setting best practices should be followed:

- When setting initial targets, use past projects with similar inputs and goals to estimate possible achievements when possible

- Minimize revisions to avoid overfitting the project's definition of success to the limitations or opportunities encountered during the project (keep revisions to no more than once/year)

- Document and justify all initial targets and subsequent revisions

Indicators themselves should also be set according to best practices, such as those established by USAID (2018) and presented below. Indicators should be:

- **Direct:** the indicator "*clearly measures the intended result.*"

- **Objective:** the indicator "is *unambiguous about 1) what is being measured and 2) what data are being collected.*"

- **Useful for management:** the indicator "*provides a meaningful measure of change over time for management decision-making.*"

- **Attributable:** the indicator "*can be plausibly associated with [the] interventions.*"

- **Practical:** the indicator "*data can be collected on a timely basis and at a reasonable cost.*"

- **Adequate:** the indicator or set of indicators is *"sufficient to measure the stated result."*

- **Disaggregated, as necessary:** indicator data are broken down by age, gender, location, or other critical aspects to aid in decision-making.

1. Strengthening poorly defined indicators

Indicators reviewed in this synthesis varied in their adherence to these best practices. In Table C.2 we present four examples of indicators drawn directly from PMPs reviewed for this synthesis that do not meet the criteria above, identify the issue with each indicator, and offer suggestions for how to improve the indicator.

Table C.2. Sample indicators

Indicator	Poorly defined indicator	Improved indicator
1	*Number and percentage of inspectors trained on the new data system*	• *Number of inspectors trained on the new system* • *Percentage of all inspectors trained on the new system*
	This indicator objective is **not direct** because "number and percentage" implies that multiple figures will be reported, and only one can be reported per item in the PMP.	These indicators are **direct** because they are each particular to one data point and they clearly measure the intended result.
2	*No. of inspectors using new data system*	*Number of inspectors who have logged onto the data system at least two times in each of the previous six months.*
	This indicator is **not objective** because the indicator does not clearly define "use". Is the figure determined by how many log-in credentials are assigned? By the number of inspectors that use the system daily?	This indicator is **objective** because "logged onto the data system at least two times" tells the M&E specialist how to assess the number of inspectors. The number should capture the inspectors who regularly access the data system each month.

Indicator	Poorly defined indicator	Improved indicator
3	*Number of new/revised procedures/tools with support of the project used by labor and fire inspectors and privately contracted monitors supporting the National Initiative*	*Number of:* • *new procedures* • *revised procedures* • *new tools* • *revised tools* *which were developed with support of the project and are used weekly by the majority of:* • *labor inspectors* • *fire inspectors* • *privately contracted monitors supporting the* • *National Initiative*
	This indicator is **not direct** because it does not clearly measure an intended result; **not objective** because it is ambiguous, **not useful for management** because it cannot be used for decision making.	If the indicator is of interest to the implementer and donor, each permutation of the bullets here should be measured separately to produce **direct**, **objective**, and **useful** indicators, and "used by" should be defined more specifically, such as "used weekly by the majority of."
4	*Number of government entities that are charged with the investigation and/or prosecution of persons or groups that commit crimes with an anti-union motive that adopt project materials as part of their internal trainings.*	*Number of national government entities that are charged with the:* • *investigation* • *prosecution* *of persons or groups that commit crimes with an anti-union motive that include project:* • *Case management system* • *Labor law toolbox* *as part of their internal:* • *onboarding* • *ongoing trainings.*
	This indicator is **not direct** because it does not clearly measure an intended result; **not objective** because it is ambiguous in measuring multiple things, **not useful for management** because its ambiguity prevents evidence-based decision making; and **not attributable** because factors greater than the project's activities likely drove the number in question.	If the indicator is of interest to the implementer and donor, each permutation of the bullets here should be measured separately to produce **direct**, **objective**, and **useful** indicators. Finally, the project should identify the ways in which changes can be attributed to project efforts. If political factors are likely to countervail the project's efforts to achieve the inclusion of these items in the entities' trainings, then the implementer should measure the contribution to agencies in another way.

2. Incorporating sustainability into indicators

Whenever possible, medium- and long-term outcomes should have indicators measuring likelihood of sustainability. Drawing from Rogers and Coates (2012), sustainability indicators should be selected to correspond with the following domains.

- Medium-term outcome domains:
    - Sustained motivation
    - Sustained resources
    - Sustained capacity

- – Sustained linkages; and
- Long-term outcome domains:
  - – Sustained service delivery
  - – Sustained access
  - – Sustained demand

In Table C.3, we provide a list of example indicators derived from PMPs reviewed for the synthesis review that meet the USAID criteria and may provide a useful basis for OTLA to draft a list of required indicators.[3] For the hypothetical projects with indicators in Table C.3, sustainability outcome domains could be represented by indicators in the medium- and long-term outcome sections.

Table C.3. Exemplary indicators

Inputs/Activities
Worker: Number of fire and building safety trainings conducted in workplaces organized in the last 6 months
Government: Number of labor complaint management protocols developed
Employer: Number of advisory visits to participating factories
Outputs
Worker: Number of female union leaders trained on reporting hazards to factory managers and GOB
Government: Number of MAST conciliators trained on labor complaint management protocols
Employer: Number of compliance assessment reports completed
Short-term outcomes
Worker: Percent participants with improved knowledge of fire/building safety and basic hazard reporting, as shown by an improvement of at least 10 percentage points between pre- and post-tests.
Government: Percent of MAST conciliators who report greater confidence in their labor complaint management skills, as shown by an improvement of at least 20 percentage points between pre- and post-surveys.
Employer: Program revenue from assessment subscriptions in the reporting period
Medium-term outcomes
Worker: Percent of worker reports resulting in remediation of hazard
Worker (sustainability-oriented indicator): Number of factories where workers form complaint-processing committees (*sustained capacity and linkages*)
Government: Percent of labor complaints received that were followed up by quarter
Government (sustainability-oriented indicator): Percent change in projected MOL allocation to inspectorate for next FY (*sustained resources*)
Employer: Average non-compliance rate of participating factories on publicly reported labor issues
Employer (sustainability-oriented indicator): Number of employers who seek ongoing technical assistance to remediate outstanding compliance problems (*sustained motivation*)
Long-term outcomes
Worker (sustainability-oriented indicator): Percent of workers in targeted factories that indicate interest in future trainings from union leadership (*sustained demand*)
Government (*sustainability-oriented indicator*): Number of inspections pre-approved for next FY (*sustained service delivery*)
Employer (sustainability-oriented indicator): Percent of assessment costs covered by international buyers' subscriptions (*sustained access and demand*)

---

[3] Common indicators could allow for easy comparison of progress across similar projects and allow reviewers (and DOL) to identify trends across similar interventions.

Note: derived from PMPs of the following projects: Bangladesh SC F&BS, Haiti MCB, Haiti BW, and Jordan BW. Medium- and long-term sustainability domains for each example sustainability indicator are shown in parentheses.

OTLA and implementers should consider **developing medium- and long-term sustainability indicators during project initiation** to (1) keep the vision for post-project impact sustainability in mind when designing RFs and PMPs and (2) keep projects accountable to their goals for sustainability during implementation.

C. Additional resources

- ILAB Grantee Resources Site:
  https://www.dol.gov/agencies/ilab/resources/grants

- FY 2019 USDOL Management Procedures and Guidelines (including for M&E):
  https://www.dol.gov/sites/dolgov/files/ILAB/OTLA_2019_06_20_MPG_2019_FINAL.pdf

- USAID Performance Monitoring Indicators:
  https://www.usaid.gov/project-starter/program-cycle/cdcs/performance-monitoring-indicators

- USAID Developing Results Frameworks:
  https://www.usaid.gov/project-starter/documents/1865/technical-note-developing-results-frameworks

- Example of measuring sustainability: Rogers and Coates (2012):
  Sustaining Development: A Synthesis of Results from a Four-Country Study of Sustainability and Exit Strategies among Development Food Assistance Projects:
  https://www.fantaproject.org/sites/default/files/resources/Exit-Strategies-Synthesis-ExecSummary-Jan2017.pdf