

ILAB Synthesis Review

Final Report

September 18, 2020

Josh Meuth Alldredge and Sarah Liuzzi

Submitted to:

United States Department of Labor
International Labor Affairs Bureau
Office of Trade and Labor Affairs
200 Constitution Ave, NW
Washington, DC 20210

Contracting Officer's Representative:
Lili Bacon

Submitted by:

Mathematica
1100 1st Street, NE, 12th Floor
Washington, DC 20002-4221
Telephone: (202) 484-4688
Facsimile: (202) 863-1763
Project Director: Sarah Liuzzi
Reference Number: 50901

DISCLAIMER

This report was prepared for the U.S. Department of Labor (USDOL) Office of Trade and Labor Affairs by Mathematica, under contract number GS-10F-0050L. The views expressed are those of the authors and should not be attributed to USDOL.

All references to non-governmental companies or organizations, their services, products or resources, in this report are offered for informational purposes and should not be construed as an endorsement by the Department of Labor, of any of the companies or organizations, or their services, products or resources.

The Department of Labor does not endorse, takes no responsibility for, and exercises no control over non-governmental organizations' Web sites mentioned in this report or its views, or contents; nor does it vouch for the accuracy of the information contained on the organizations' websites.

ACKNOWLEDGMENTS

The authors acknowledge several people whose contributions have greatly enhanced the quality of this report. Lili Bacon has offered guidance and insight throughout the preparation and execution of the review that have enhanced our understanding of OTLA's work. Margaret Scotti and Lauren Jowell also provided insightful feedback on our methodology that shaped our approach.

At Mathematica, we acknowledge the contributions of Sarah Dolfin and Nan Maxwell, who both reviewed and provided helpful feedback on the methodology plan. Sarah Dolfin also provided helpful feedback on this report. Gwyneth Olson provided graphic design assistance for infographics. Dorothy Bellow and Colleen Fitts helped prepare the report. Naomi Dorsey provided skillful project management.

This page has been left blank for double sided copying.

Contents

ACKNOWLEDGMENTS.....	iii
Acronyms.....	xi
Executive Summary	xiii
I. Introduction	1
II. Research questions and analytic approach.....	2
A. Research questions	2
B. Analytic approach.....	2
C. Limitations and strategies to overcome limitations.....	4
III. Findings	6
A. Project characteristics	6
1. Project locations.....	8
2. Target stakeholders.....	9
3. Funding and period of performance.....	9
4. Project objectives and target industries	11
5. Principal inputs and activities	11
6. Implementers	13
B. Evaluation reports and data available for the synthesis review.....	13
1. General evaluation quality	14
2. Typical evaluation approach.....	15
3. Qualitative data sources.....	15
4. Quantitative data sources.....	16
C. Factors associated with project effectiveness.....	20
1. Factors OTLA can influence	21
2. Factors under implementer control	27
3. Contextual factors	33
4. The role of government capacity and political will, climate, and instability	33
5. Projects operating in settings where workers are not free to organize	34
D. Challenges and solutions	35
1. Description of common challenges.....	35
2. The most common challenges and solutions applied.....	37

- IV. Conclusions and recommendations42
 - A. Recommendations from evaluations42
 - B. Conclusions and recommendations derived from synthesis of evaluations.....43
 - 1. Conclusions and recommendations for project priorities and design factors related to project effectiveness.....43
 - 2. Recommendations related to common challenges and strongest solutions45
 - 3. Recommendations related to evaluations and data quality46
- Appendix A: List of documents used A.1
- Appendix B: Detailed description of analysis approach B.1
- Appendix C: Theory of change and indicator guidance C.1

Tables and boxes

Table 1. The five most commonly observed challenges and solutions applied	xvi
Table 2. Evaluator-provided recommendations	xvii
Table 3. Conclusions and recommendations for project priorities and design factors related to project effectiveness.....	xviii
Table 4. Conclusions and recommendations related to challenges and solutions	xix
Table 5. Conclusions and recommendations related to evaluations and data quality	xx
Table II.1. Limitations to the synthesis review and strategies to overcome them	4
Table III.1. Characteristics of projects selected for the synthesis review	7
Box III.1. Electronic case management systems (ECMS).....	12
Box III.2: Project effectiveness and sustainability scoring system.....	21
Table III.2. General trends in project factors and achievement by primary target group	23
Box III.3. Better Work programs	25
Box III.4. Projects with gender equality as objectives or impacts.....	25
Box III.5 Projects engaging vulnerable populations	27
Table III.3. Challenges and solutions	36
Table III.4. Example challenge and solution: Limited government capacity (includes poor systems, turnover, and corruption).....	37
Table III.5. Example challenges and solutions: political will, political resistance, and low government ownership of project activities.....	37
Table III.6. Example challenges and solutions: political uncertainty, unrest, and turmoil; rapidly changing conditions	38
Table III.7. Example challenges and solutions: Insufficient implementer capacity	39
Table III.8. Example challenges and solutions: Low union or labor federation capacity.....	39
Table VI.1 Evaluator-provided recommendations.....	42
Table A.1. Projects and evaluations included in the synthesis review	A.3
Table C.2. Sample indicators	C.6
Table C.3. Exemplary indicators	C.8

This page has been left blank for double sided copying.

Figures

Figure II.1. Logic model framework to analyze projects in the synthesis review	2
Figure III.1. Number of projects included in the synthesis review, by country	9
Figure III.2. Project funding	10
Figure III.3 Project periods of performance	10
Figure B.1. Screenshot of analytical table	B.6

This page has been left blank for double sided copying.

Acronyms

BC-PLIP	Building the Capacity of the Philippines Labor Inspectorate (project in the Philippines)
BUCCPEP	Building Union Capacity to Reduce Precarious Employment in Peru (project in Peru)
BW	Better Work
BWB	Better Work Bangladesh
BWH	Better Work Haiti
BWJ	Better Work Jordan
BWV-UCD	Better Work Vietnam – Union Capacity Development
CAL	Workers’ Rights Centers for the acronym in Spanish (project in Colombia)
ECMS	Electronic case management system
EEFMP	Egypt Export Factory Monitoring Project (project in Egypt)
F&BS	Fire and building safety
GTU	Grassroots trade union
ICLLG	Improved Compliance with Labor Laws in Georgia (project in Georgia)
IGUALDAD	Initiative to Guard Against Labor Discrimination (project in Mexico)
ILAB	International Labor Affairs Bureau
ILO	International Labor Organization
ILO F&BS	International Labor Organization Fire and Building Safety (project in Bangladesh)
ILO-IR	International Labor Organization-Industrial Relations (project in Vietnam)
ILS	International labor standards
LAC	Latin America and the Caribbean
LL-IRR	Support for Labor Law and Industrial Relations Reform in Malaysia (project in Malaysia)
MCB	Building Capacities of the MAST to Ensure Labor Law Compliance in the Haitian Apparel Sector (project in Haiti)
MENA	Middle East and North Africa
OCFT	Office of Child Labor, Forced Labor, and Human Trafficking
OTLA	Office of Trade and Labor Affairs
PCILSC	Promoting Compliance with International Labor Standards in Colombia (project in Colombia)
PLIP	Building the Capacity of the Peruvian Labor Inspectorate (project in Peru)

PMP	Performance monitoring plan
RF	Results framework
RGR	Results-to-goals ratio
RMG	Ready-made garment
SC	The Solidarity Center (affiliated with AFL-CIO)
SC F&BS	Solidarity Center Fire and Building Safety (project in Bangladesh)
SWOG	Strengthening Workers' Organizations in Georgia (project in Georgia)
USDOL	United States Department of Labor
WRC	Workers' rights centers

Executive Summary

The Office of Trade and Labor Affairs (OTLA) in the Bureau of International Labor Affairs (ILAB) within the U.S. Department of Labor (USDOL) works to improve labor law enforcement to protect workers' rights and promote a fair global playing field for workers in the United States and around the world. As part of this effort, OTLA manages and funds technical assistance projects designed to improve workers' rights and working conditions and ensure compliance with labor laws. The projects are diverse in their strategies and implementing environments, but all aim to improve the capacity of governments, workers, and or employers to enforce and improve labor protections. Projects may target one, two, or all three stakeholder groups in what are known as tripartite projects.

OTLA has invested in performance evaluations of the technical assistance projects it funds. OTLA now seeks to identify trends and patterns in the findings presented in the performance evaluations of 19 projects from 12 countries designed to improve labor law enforcement and compliance. To that end, OTLA contracted with Mathematica in September 2019 to conduct the ILAB Synthesis Review.

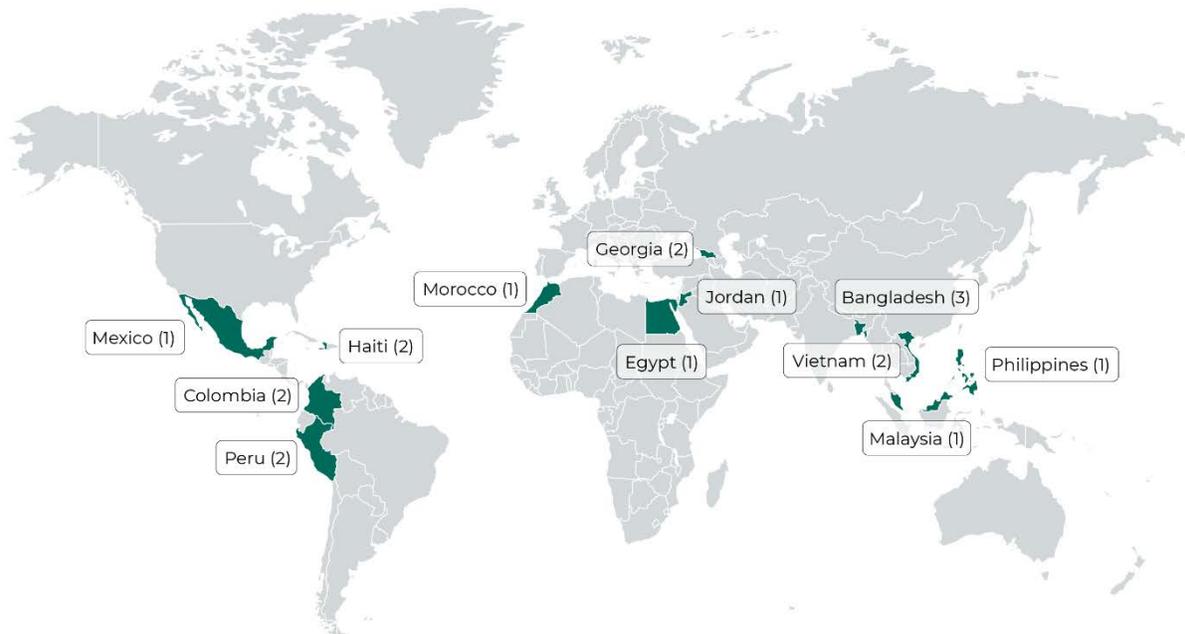
The overall research objective of the synthesis review is to generate information and recommendations that will support ILAB in deciding (1) which future projects to fund to improve labor law enforcement and compliance around the world, (2) the structure and support needed for those projects, and (3) how to maximize the effectiveness of the funding investments made. We have developed a set of research questions in support of this objective, as well as to respond to our overarching question: How should OTLA use its resources moving forward to maximize its effectiveness in improving labor law enforcement and compliance around the world? We have organized our findings into four parts:

1. Description of the projects included in the synthesis review and data available for the review
2. Factors associated with project effectiveness
3. Challenges projects face and most effective solutions
4. Conclusions and recommendations

Our review was based on information found in external evaluation reports and monitoring and evaluation data for the 19 projects selected for this review, which we coded into categorical variables. We then scored projects by effectiveness, identified relationships between project factors and effectiveness, and examined project challenges and solutions discussed by evaluators.

A. Description of the projects included in the synthesis review and data available for the review

The projects selected for the review were diverse in geography, scope, sector, and strategy. The 19 projects selected for review span 12 countries and 5 regions (see Figure 1). Implementers typically conducted project activities with national-level stakeholders, such as Ministries of Labor, employers' associations, or unions, and/or with local-level stakeholders, such as individual factories and workers within them. Project aims ranged from building government inspection and enforcement capacity, to supporting employer compliance, and or building worker and union capacity for advocacy. The most common target sector was the ready-made garment (RMG) industry, and the most common project inputs were training and technical assistance. Project funding ranged from \$0.9 to \$13.3 million and periods of performance ranged from 2 to almost 11 years. The International Labor Organization (ILO) and the Solidarity Center (SC) were the primary implementers in 15 of the 19 projects.

Figure 1. Number of projects included in the synthesis review, by country

Note: The numbers in parentheses refer to the number of projects reviewed in each country.

The review was based on evaluation reports prepared by external evaluators and implementers' monitoring evaluation data, both of which varied in quality and completeness. The evaluation reports used for this synthesis were medium-quality and varied substantially in the level of detail provided. All the evaluations reviewed were mixed methods, non-causal performance evaluations. Evaluations relied on qualitative data collected through document reviews, key informant interviews, or focus group discussions. We also drew from projects' performance monitoring plans (PMPs), which compare project results to targets set for key indicators of inputs, outputs, and short-, medium-, and long-term outcomes.

Some aspects of our review are limited by the quality and scope of the reviews and PMP data. The evaluation reports offered rich information on the projects, but varied in quality and did not always offer complete information. Only half offered information on gender-related outcomes and many PMPs were missing information on key indicators. When reports were published before projects ended, the authors were unable to comment on final outcomes and had limited capacity to comment on the likelihood that projects' impacts would be sustained.

B. Factors associated with project effectiveness

For each project, we reviewed over 200 project characteristics to identify factors that may be associated with project effectiveness. We have organized the factors we reviewed into factors OTLA can influence, such as project size or scope; factors under implementer control, such as project management approach; and contextual factors. Our findings are based on assessing characteristics of projects that we evaluated as most effective in meeting their goals as compared to those that were less effective and identifying factors that may relate to effectiveness.



Factors that OTLA can influence

- **Projects with longer periods of performance and larger budgets were among the most effective.** This result may have been driven by projects benefiting from having sufficient time to carry out meaningful project activities. Projects with larger budgets, which is driven by project period of performance, were also among the more effective.
- **Projects primarily targeting government agencies, such as ministries of labor, were among the less effective projects.** Projects that primarily targeted employers, workers, or were fully tripartite, were more effective.
- **Better Work (BW) projects were moderately effective in achieving their goals,** and the most successful inputs of those programs were 1) training and technical assistance with factory managers and 2) factory-level committees with workers and factory managers.



Factors under implementer control

- **Not surprisingly, projects that were well-managed and well-designed showed greater effectiveness.** Projects with high-quality management and high efficiency were more likely to achieve their stated goals. Similarly, projects with a well thought-out and articulated theory of change were more effective than projects with theories of change that were less coherent or not evidence-based. Poorly managed projects, including those that suffered from implementation delays, were less effective.
- **Projects that were explicitly tripartite or explicitly not tripartite did better than projects that made limited commitments to a tripartite approach.** Tripartite projects are complex and this result suggests that implementers position themselves for success by planning for the complexities of a tripartite project from the outset.
- **Projects that were well positioned for the unexpected were more effective.** These included projects that anticipated and prepared for potential risks and that were flexible in their implementation.



Contextual factors

- **We identified few contextual factors that appeared to have a strong, direct relationship with project effectiveness.** This result may be driven by projects’ ability to develop effective strategies to overcome challenges, or by the fact that other factors simply had more dominant impacts on effectiveness.
- **Although limited government capacity and political will represented recurring challenges for projects, these factors did not appear to have strong relationships with project effectiveness.** Intervening factors—such as the quality of project management or the success of solutions applied to the challenges—may have had more immediate effects on project progress than the presence of the economic, social, political, or government factors themselves.
- **Of the evaluation reports that cited contexts where workers were not free to organize, all of those projects devised strategies to be effective in spite of that challenge.** They addressed that issue through policy reform supports, union capacity development, and contracting guidance with moderate to high success.

C. Challenges and solutions

Across the 19 projects, the evaluations reported challenges 73 times (most of which were observed in more than one project), and reported solutions implemented to mitigate those challenges 26 times. We have summarized the five most common challenges projects faced and solutions used in Table 1. The report includes full results including less commonly observed challenges.

Table 1. The five most commonly observed challenges and solutions applied

Challenges	Times occurred	Solutions	Times used	Success 0-3
Limited government capacity	10	Provide assessment services to employers when Ministry of Labor inspections are not performed	1	2
Limited political will or political resistance	9	Narrowing project activities to focus on most successful technical aspects	3	2
		Engage more with partners	1	2
		Be flexible with legal and policy reform proposals	1	0
		Deploy project personnel to engage Ministry of Labor and seek MOU to promote sustainability	1	0
Political uncertainty, unrest, and turmoil; rapidly changing conditions:	8	Be flexible with activities, focus on communication, and build partner human resources	3	2
Insufficient implementer capacity, including with monitoring	7	Bring in consultants	1	2
		Simplify indicators and improve measurement strategy	1	1
Low union or labor federation capacity (including workers lacking collective bargaining tools)	5	Co-conduct a strategic planning exercise	1	2

Note: “times occurred” refers to the number of times the challenge occurred (not the number of projects in which it occurred). “Times used” refers to the number of times that the solution was deployed for the given challenge, and “success 0-3” indicates the score we assigned to the solution in terms of its efficacy in mitigating the challenge, with 0 indicating no mitigation and 3 indicating complete mitigation.

D. Conclusions and recommendations

This report synthesized the findings of evaluations of 19 projects that aimed to build government inspection and enforcement capacity, support employer compliance with labor laws, and or build worker and union capacity for advocacy. Although the nature of this review precludes us from identifying causal relationships between project characteristics and outcomes, our review revealed relationships between some project features and project effectiveness, recurring challenges that OTLA and implementers should prepare for, and opportunities for improving project data practices moving forward. We begin this section by presenting recommendations made in the evaluation reports before presenting our own conclusions and recommendations.

Evaluation reports all made recommendations and the most common recommendation was to combine, expand, or realign project activities to better fit needs. Each of the 19 evaluations presented recommendations for improving the performance, effectiveness, and sustainability of the project at hand (if ongoing) or future projects. We consolidated the numerous suggestions into 11 main recommendation groups in the Table 2. Most of the recommendations offered by evaluators align with our recommendations.

Table 2. Evaluator-provided recommendations

Recommendation group	Number of projects where recommended
Combine, expand, or realign project activities to better fit needs ¹	17
Improve M&E systems, project PMP, and/or accountability requirements	14
Provide greater assistance and resources to low-capacity government partners	13
Revise sustainability plan or embed sustainability into project design	11
Help partners engage one another more, open greater communication	8
Clarify project priorities, structure, and roles for staff and partners	6
Adjust budget, allocate funds more clearly, or set up outcome-based budgeting	4
Increase funding / streamline disbursement of funds (particularly in ILO projects)	3
Provide greater assistance to unions or employers	3
Improve gender-conscious programming	3
Showcase or disseminate project successes	2

Our recommendations are based on a thorough synthesis review of the 19 evaluation reports, but our findings described above and summarized below are subject to some limitations. These conclusions and recommendations represent our assessment of patterns and trends in projects’ level of effectiveness in achieving inputs, outputs, and outcomes and factors that may influence their effectiveness as well as our qualitative analysis of evaluators’ comments on project implementation. The trends we have identified may be based on real relationships; however, given that 1) our analysis is based on a finite sample of evaluations of 19 projects and 2) lower-quality evaluations tended to report higher project effectiveness, the relationships we have described in this report may instead be due to chance or overstated findings. Nonetheless, we consider these findings worth taking into consideration.

¹ This group includes recommendations such as “Combine Fire and Building Safety and OSH” (*combine*) or “Expand EEFMP to all governorates” (*expand*) or “Ensure that project priorities are clearly reflected in the project’s logic model so that the project monitoring plan (PMP) and the work plans better reflect the project’s goals” (*realign*). Each recommendation in this group relates to delivering interventions that are comprehensive, right-sized, and linked to core goals and beneficiary needs.

We present conclusions and recommendations in three areas: project priorities and design factors related to project effectiveness, most common challenges and solutions, and recommendations related to evaluations and data quality.

Conclusions and recommendations for project priorities and design factors related to project effectiveness. Table 3 presents key findings on project characteristics found among the most effective projects and recommendations based on these findings.

Table 3. Conclusions and recommendations for project priorities and design factors related to project effectiveness

Conclusions	Recommendations
Project design	
<ul style="list-style-type: none"> Projects with more comprehensive scopes appear to be more effective² Projects with larger budgets, longer periods of performance, and fully tripartite approaches were more likely to meet their goals than shorter projects with smaller scope. Projects aligned with ongoing activities in the target country had more success. 	<ul style="list-style-type: none"> USDOL: Focus OTLA investments in more comprehensive, flexible, long-term projects that are aligned with ongoing activities in the target country.
<ul style="list-style-type: none"> Projects targeting workers, unions, and labor federations tended to have greater effectiveness in meeting project goals than those targeting government or employers 	<ul style="list-style-type: none"> USDOL: Increase the share of OTLA projects with workers, unions, or federations as primary project target
<ul style="list-style-type: none"> Projects with evidence-based theories of change and those with high-quality management structures were more likely to meet their goals and overcome challenges. 	<ul style="list-style-type: none"> USDOL: Assess evidence base for project before the solicitation stage. Implementer: Review evidence base for project immediately after award.
<ul style="list-style-type: none"> Projects were more likely to improve gender equality outcomes when gender was part of project design. 	<ul style="list-style-type: none"> USDOL: Consider requiring projects to mainstream gender considerations into their designs—universally or in a targeted manner
Project context	
<ul style="list-style-type: none"> Contextual factors like economic or political conditions did not have a strong, direct relationship with project success, but this is likely to be because projects used strategies to manage conditions. 	<ul style="list-style-type: none"> USDOL: Provide implementers with a toolkit with strategies to handle common challenges (following guidance in Section III.D of this report).
<ul style="list-style-type: none"> The most common types of challenges involved limited government capacity and low political will. 	<ul style="list-style-type: none"> USDOL: Provide implementers with a toolkit with strategies to handle common challenges (following guidance in Section III.D of this report). Implementers: Offer governments training and technical assistance.

² Projects that had overly ambitious goals in their PMPs (ones that were not calibrated to what was feasible given time and budget) indeed fell short of them. However, projects that were ambitious in their scope, as in ambitious in their partnership building or in their comprehensive delivery of services to a given sector, tended to do better in meeting their goals. Projects that sought to approach *multiple actors and aspects* of a given sector were more successful than those who were more focused on a single actor or aspect of the sector.

Conclusions	Recommendations
Project implementation	
<ul style="list-style-type: none"> • Implementer capacity plays a key role in project effectiveness. 	<ul style="list-style-type: none"> • USDOL: Select implementers with demonstrated capacity to implement projects or be prepared to provide initial and ongoing guidance and support for implementers.
<ul style="list-style-type: none"> • Weak monitoring and evaluation systems and PMPs hindered projects in achieving, tracking, and reporting their progress. • Projects with stronger monitoring and evaluation systems and PMPs could identify specific areas of success and concern and adjust strategies as needed. 	<ul style="list-style-type: none"> • USDOL: Require strong monitoring and evaluation, PMP, and accountability systems and support implementers in these efforts with written resources and mentoring.
<ul style="list-style-type: none"> • Projects' strategies have supported vulnerable populations; however, only three projects explicitly targeted vulnerable populations such as migrants or illegally contracted workers. 	<ul style="list-style-type: none"> • USDOL: Consider incorporating more services for vulnerable populations into project designs.

2. Conclusions and recommendations related to common challenges and strongest solutions

Challenges confronted every project, and implementers applied solutions with mixed success. By drawing lessons from successful and unsuccessful solutions, we can offer recommendations for strategies that could help future projects overcome similar challenges.

Table 4. Conclusions and recommendations related to challenges and solutions

Conclusions	Recommendations
<ul style="list-style-type: none"> • Some challenges were common, including limited government capacity, low political will or resistance to project activities, political uncertainty or social unrest, insufficient implementer capacity, and low union or labor federation capacity. 	<ul style="list-style-type: none"> • USDOL: Develop and share a toolkit with resources for implementers to draw from to manage common challenges. The guidance in the toolkit could draw from the findings on most effective solutions, as described in Table 1. • USDOL: Use the following tools during project conception and design: <ul style="list-style-type: none"> – Needs and opportunities assessments – Risk analysis systems – Stakeholder analysis tools

3. Conclusions and recommendations related to evaluations and data quality

Evaluations reviewed for this report were mixed-methods, non-causal performance evaluations, and were generally medium-quality with opportunities for improvement. Most of the qualitative data sources were subject to bias while the quantitative data provided through implementers' PMPs were not consistently reported and targets for indicators were at times adjusted too often to be meaningful. The validity of evaluation findings was also threatened by the timing of the evaluations because some evaluations were published before projects ended. The following recommendations are designed to help USDOL, evaluators, and implementers avoid those evaluation and data quality deficiencies. We recommend that USDOL emphasize clear expectations on data reporting requirements expected of implementers and provide guidance to optimize the quality of data collected and used for projects' evaluations.

Table 5. Conclusions and recommendations related to evaluations and data quality

Conclusions	Recommendations
Evaluation and data collection timing	
<ul style="list-style-type: none"> Evaluation methods were not rigorous. In many cases, more rigorous methods were not feasible, but considering evaluation options before project start may facilitate the use of rigorous methods in the future. 	<ul style="list-style-type: none"> USDOL: Engage evaluators at the project design stage (preferably pre-award) to support the development of the most rigorous evaluation design possible.
<ul style="list-style-type: none"> Data collected did not always reflect full project implementation 	<ul style="list-style-type: none"> USDOL: Require evaluations to collect data at project end—not before—to allow for clearest view of project success. Consider allowing for data collection one year after project completion to assess longer-term outcomes and sustainability. USDOL: Provide evaluators with the resources and time to conduct fieldwork that captures a full view of the project.
Data quality and completeness	
<ul style="list-style-type: none"> Some items in projects' results frameworks are not included in PMPs. 	<ul style="list-style-type: none"> USDOL: Require and verify that projects' PMPs include indicators (and targets) for outputs and all short-, medium-, and long-term outcomes.
<ul style="list-style-type: none"> PMP data are incomplete. 	<ul style="list-style-type: none"> USDOL: Require and verify that projects consistently report indicators in their PMPs. USDOL: Provide an easy-to-use electronic platform for implementers to report PMP data.
<ul style="list-style-type: none"> PMP targets are of limited utility. 	<ul style="list-style-type: none"> USDOL: Provide guidance to implementers on appropriate target-setting approaches (use prior projects' results when possible to set targets, limit frequency of updating targets, and document rationale for changing targets). Implementers: Provide written justification for all revisions to targets.
<ul style="list-style-type: none"> Indicators varied by project, limiting the potential for cross-project comparisons. 	<ul style="list-style-type: none"> USDOL: Consider identifying a small set of indicators that would be required across projects or across projects with similar long-term objectives.
<ul style="list-style-type: none"> PMPs often lacked data on gender outcomes or disaggregation by gender. 	<ul style="list-style-type: none"> USDOL: Encourage or require implementers to report on gender outcomes and data disaggregated by gender. Implementers: Report gender outcomes and data disaggregated by gender.
<ul style="list-style-type: none"> Qualitative data were subject to bias, including sample selection bias, recall bias, and results that are not representative because of small sample sizes. Furthermore, if interviews or focus groups are conducted within earshot of workers' supervisors, they may not provide accurate or full information. 	<ul style="list-style-type: none"> USDOL: Support evaluators in developing a representative sampling frame of potential interviewees from which to draw the sample for qualitative data collection. USDOL: Use rapid scorecards or small surveys to discreetly gather consistent data from all participants in qualitative data collection.
Evaluation report completeness	
<ul style="list-style-type: none"> Evaluation reports do not always report or critically assess implementers' theories of change or results frameworks. 	<ul style="list-style-type: none"> USDOL: Require evaluators to present and discuss projects' results frameworks and theories of change and accompanying assumptions in their evaluations.
<ul style="list-style-type: none"> Few evaluation reports reported on gender. 	<ul style="list-style-type: none"> USDOL: Require evaluators to assess projects' gender approaches and impacts.
<ul style="list-style-type: none"> Evaluators did not always conduct validation workshops with stakeholders to review findings, as required by USDOL. 	<ul style="list-style-type: none"> USDOL: Ensure that evaluators plan for and execute validation workshops.
<ul style="list-style-type: none"> Evaluators did not consistently have access to project budgets and financial data. 	<ul style="list-style-type: none"> Require that implementers make financial data available to evaluators.

I. Introduction

The Office of Trade and Labor Affairs (OTLA) in the Bureau of International Labor Affairs (ILAB) within the U.S. Department of Labor (USDOL) works to improve labor law enforcement to protect workers' rights and promote a fair global playing field for workers in the United States and around the world. As part of this effort, OTLA manages and funds technical assistance projects designed to improve workers' rights and working conditions and ensure compliance with labor laws. The division has over 12 active technical assistance projects in 14 countries and has supported technical assistance projects in over 72 countries since 1995. The projects are diverse in their strategies and implementing environments, but all aim to improve the capacity of governments, workers, and or employers to enforce and improve labor protections. Projects may target one, two, or all three stakeholder groups in what are known as tripartite projects.

OTLA has invested in performance evaluations of the technical assistance projects it funds and seeks to identify trends and patterns in the findings presented in performance evaluations of 19 projects designed to improve labor law enforcement. To that end, OTLA contracted with Mathematica in September 2019 to conduct the ILAB Synthesis Review. This synthesis review:

- Identifies trends in findings, lessons learned about project features and implementation strategies that appear to support successful outcomes, and recommendations for strategies to prioritize moving forward;
- Provides insights on the theories of change that underlie the projects, types of interventions, and promising strategies for bolstering workers' rights around the world; and
- Communicates findings, conclusions, and recommendations from the synthesis review.

The 19 projects selected for this review are diverse geographically and in terms of budget and specific focus. We describe the projects in more detail in section III.A. Appendix A provides basic information on each of the evaluation reports included in the synthesis review.

The intended audience for this report includes decision makers within USDOL, donors deciding which interventions to fund, and governments and organizations implementing strategies to improve working conditions.

This page has been left blank for double sided copying.

II. Research questions and analytic approach

A. Research questions

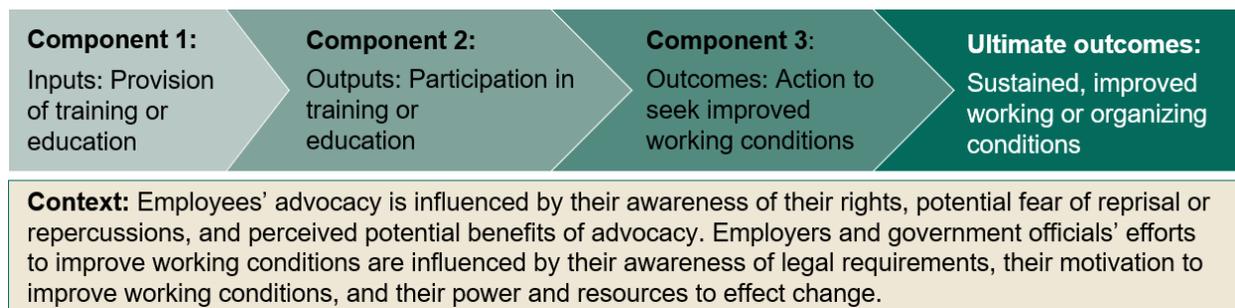
The overall research objective of the synthesis review is to generate information and recommendations that will support ILAB in deciding (1) which future projects to fund to improve labor law enforcement and compliance around the world, (2) the structure and support needed for those projects, and (3) how to maximize the effectiveness of the funding investments made. We have developed a set of research questions in support of this objective, as well as our overarching question: How should OTLA use its resources moving forward to maximize its effectiveness in improving labor law enforcement and compliance around the world? The following sub-questions will structure our analysis of the evaluation reports:

1. What are the characteristics of the projects that were evaluated?
2. How were the evaluations conducted?
3. What were the evaluations' findings on implementation, outputs, and outcomes?
4. What challenges did projects typically face?
5. What strategies were used to address those challenges, and which strategies appeared to be most effective?
6. What trends and patterns emerged from the evaluations' findings?

B. Analytic approach

In this section, we summarize the steps we took in analysis of the 19 projects.³ Our analysis is premised on a logic model Mathematica developed as a framework to understand the projects, shown in Figure II.1. This logic model shows the progression of an example capacity-building project from the provision of inputs to the production of outputs and achievement of immediate and ultimate outcomes.⁴ The model also includes projects' context because of its role in facilitating or limiting a project's effectiveness.

Figure II.1. Logic model framework to analyze projects in the synthesis review



³ Appendix B provides a detailed description of our analysis approach.

⁴ OTLA now uses the terms outputs and short-term, medium-term, and long-term outcomes to convey theories of change (TOCs) in results frameworks. However, the projects reviewed for this synthesis used a variety of labels to convey their TOCs. We capture these TOCs using the following four labels: inputs - execution of planned activities; outputs - evidence of planned activities (participation in this example); outcomes - application of new skills or capacities, and ultimate outcomes - achievement of major project goals.

With this logic model in mind, we extracted information from evaluations, coded it to categorical variables, scored projects by effectiveness, identified relationships between project factors and effectiveness, and examined project challenges and solutions discussed by evaluators. We present each step below.

- **Extracting information.** We pulled detailed information from each of the 19 evaluations into a 61-item rubric that covered project characteristics (including planned inputs, outputs, outcomes and ultimate outcomes), evaluation features, project implementation and actual outcomes, contextual factors, and project challenges and solutions.
- **Coding information.** We broke down each rubric item into multiple binary or categorical variables and coded each of the details on each project to the variables. For example, for the rubric item *Factors or opportunities that will contribute to project success*, we produced four columns, each headed by a supporting factor group, such as alignment with other projects or efforts in the country, and gave each project a “1” if it had that supporting factor and a “0” if it did not have it. This coding process, together with the addition of composite and summary variables, produced 238 variables and 4,522 values.
- **Scoring effectiveness.** To assess the effectiveness of each project in delivering inputs, producing outputs, and achieving immediate and ultimate outcomes⁵ as planned, we comprehensively examined projects as described in the evaluations and used an A-B-C scoring system (please see Appendix B for details on the scoring approach). We also assessed and scored each project using an A-B-C system for the likelihood that impacts would be sustained beyond the duration of the grant funding.
- **Identifying relationships.** To analyze the relationships between project factors and A-B-C scored items, we examined trends and patterns that became visibly apparent in tables sorted to isolate projects with specific characteristics. For example, sorting by principal project targets (employer, ministry, tripartite, or worker) allows for the identification of patterns among specific target groups in terms of project management and efficiency, level of partner buy-in, and the effectiveness of projects in achieving inputs, outputs, and ultimate outcomes. After identifying associations, we returned to the original detail-rich rubric to extract examples of the relationships between variables and flesh out the relationships in our narrative of the findings.
- **Examining challenges and solutions.** Using the rubric, we found projects faced 73 unique challenges. We grouped into 21 broader challenges, such as *project team hiring difficulties and turnover*. We then examined the strategies applied by projects to each challenge and scored the success of the solution in mitigating the challenge from 0 (not successful) to 3 (completely successful).

This analytical approach allowed us to exploit the richness of the data, isolate relationships among characteristics and ultimate outcomes, and draw conclusions about what factors might support project success.

⁵ We excluded immediate outcomes from this scoring and focus on inputs, outputs, and ultimate outcomes because we care most about whether inputs were delivered as planned, outputs were produced as expected, and ultimate outcomes were achieved as desired. The degree to which immediate outcomes were reached as planned is captured through our assessment of outputs and ultimate outcomes. In some cases, PMPs did not include data on ultimate outcomes, but we were able to assess ultimate outcomes based on narrative descriptions in the evaluation reports.

C. Limitations and strategies to overcome limitations

In this review, we have identified potential relationships between project characteristics and levels of effectiveness, as well as challenges that are commonly faced and strategies used to mitigate the effects of those challenges. However, like all research, this synthesis review has limitations that affect the conclusions we are able to draw. We have summarized these limitations as well as strategies we have employed to mitigate the impacts of these limitations, when possible, in Table II.1.

Table II.1. Limitations to the synthesis review and strategies to overcome them

Limitation	Strategy
<p>We can identify relationships, but not causal impacts. As we described in the previous section and in Appendix B, much of our analysis is based on assessing variation in project effectiveness across groups of projects with different characteristics. This analysis allows us to identify potential correlations between project characteristics and outcomes, but not causal effects of those characteristics. For example, as we describe in Section III.C., among the 19 projects selected for this review, we found that the projects with longer periods of performance also had better project management. It could be that projects with longer periods of performance have stronger project management because of having time to hone those skills. However, it could also be that implementers with stronger project management skills are selected intentionally for longer-term projects. We are unable to determine whether such a relationship is causal or the direction of causality with this set of evaluations.</p>	<p>In some cases, qualitative information from the reports shed light on what might be driving project outcomes. Although we are unable to identify causal relationships, some evaluation reports include detailed descriptions of how projects were implemented. The reports may shed light on why some projects were well managed and what management challenges other projects faced. These stories may help us understand which factors are more or less likely to be important in determining project outcomes.</p>
<p>Our analysis is limited to the content of the evaluators' reports and implementers' performance monitoring plans (PMPs), which vary in quality and are subject to bias. As we describe in Section III.B, the evaluations and PMPs for the projects included in this synthesis review vary in quality from low to high quality and we know less about the projects with evaluation reports and PMPs that are lower quality. Furthermore, even the high-quality reports and PMPs are subject to the interpretation of their authors. Individual authors' expectations and prior experiences may lead them to highlight different elements of the projects, or to interpret different project outcomes in different ways. Similarly, as we discuss in Section III.B, an important part of the PMPs is the targeting process that establishes targets for projects' specific inputs, outputs, and outcomes. Implementers take different approaches to both target setting and target revising. Finally, the evaluations report on numerous factors, but some important factors may not have been reported on.</p>	<p>We focus our analysis on data believed to be more reliable. We limited our analysis to the information available in the evaluation reports and PMPs, but in recognition of the varied quality of the information, we focus our analysis on the data we consider more reliable. For example, because the process for setting indicator targets was inconsistent, we do not rely heavily on the relationship between project outcomes and targets.</p>
<p>Our own determinations of how to group projects with similar characteristics or outcomes is subjective; other reviewers may have reached different conclusions. Our analysis was conducted by a team of two and the review was done largely by one person.</p>	<p>Strategy: Whenever possible, we created rules or cut points for grouping similar groups in an effort to assess projects consistently.</p>
<p>This review's findings are based exclusively on the experiences of the 19 projects selected for the review and may not generalize to the full set of OTLA-funded projects. Our analysis is limited to the 19 projects included in the review, but much of the analysis focuses more narrowly on subgroups of projects, such as projects with a core focus on workers, which represent a subset of 6 projects.</p>	<p>Strategy: We keep the limited external validity of this review (applicability to other projects) in mind when interpreting our findings and encourage the reader to do the same.</p>

Limitation	Strategy
<p>Many evaluations do not have information on gender outcomes, limiting our ability to comment on gender aspects of the projects. As we discuss in III.C, only 10 of the evaluations examined gender issues and only 9 of the 19 projects reported gender-disaggregated indicators. Although projects' omission of gender data suggests that gender was not a priority for those projects, we are unable to fully assess how those projects handled gender without indicators disaggregated by gender.</p>	<p>Strategy: We have made observations based on the evaluations that do feature information on gender and clarify in this report the limitations to our ability to draw conclusions for all 19 projects.</p>

III. Findings

In this section, we summarize our findings in response to all six research questions. We begin in Section A by describing the characteristics of the projects selected for the synthesis review (Research Question 1) and continue in Section B with a description of the evaluation reports and data we used to conduct the review (Research Question 2). In Section C, we present the factors we identified as being associated with high or low levels of project effectiveness (Research Questions 3 and 6). Finally, in Section D, we review the challenges projects faced and the solutions that were most effective in mitigating those challenges. Each section begins with a summary of key points and concludes with specific recommendations derived from our analysis.

A. Project characteristics

Section summary: Project characteristics

- The 19 projects selected for review span 12 countries and 5 regions. Implementers typically conducted project activities with national-level stakeholders, such as Ministries of Labor, employers' associations, or unions, and/or with local-level stakeholders, such as individual factories and workers within them.
- Project funding ranged from \$0.9 to \$13.3 million and periods of performance ranged from 2 to almost 11 years.
- Projects aimed to build government inspection and enforcement capacity, support employer compliance, and or build worker and union capacity for advocacy.
- The most common target sector was the ready-made garment (RMG) industry, and the most common project inputs were training and technical assistance.
- The International Labor Organization (ILO) and the Solidarity Center were the primary implementers in 15 of the 19 projects.

Table III.1 provides basic information on the 19 projects⁶ selected for the synthesis review and the following section describes the characteristics of these projects, specifically: geography, target stakeholders, funding, periods of performance, project objectives and target industries, project inputs and activities, and implementers.

⁶ The numbering in the table follows the original numbering used in the contract. It is not continuous, because two evaluation reports were removed when DOL established the projects they evaluated were outside the scope of the synthesis review or already covered in another evaluation report document.

Table III.1. Characteristics of projects selected for the synthesis review

Country and project acronym	Region	Focus	Main activities	Primary targets ¹			Length, start and end dates	Budget (millions)
				Employers	Workers	Government		
Bangladesh (BWB)	South Asia	Labor-law compliance	Advisory services, training, compliance assessments	■	□	□	2 years, 3 months: 9/16/14–12/31/16	\$1.22
Bangladesh (SC F&BS)	South Asia	Fire and building safety	Training and technical assistance	□	■	□	5 years, 6 months: 9/30/13–3/31/19	\$1.2
Bangladesh (ILO F&BS)	South Asia	Fire and building safety	Training, regulation harmonization	□	■	■	3 years, 2 months: 11/8/13–1/31/17	\$1.5
Colombia (CAL)	LAC	Protecting workers' rights	Legal advice and supports, training	■	■	■	4 years: 12/1/12–12/1/16	\$1.8
Colombia (PCILSC)	LAC	Institutional capacity to protect workers' rights	Technical assistance and training	■	■	■	5 years, 6 months: 1/1/12–6/30/17	\$10.3
Egypt (EEFMP)	MENA	Inspection capacity and working conditions	Technical support and training, dialogue support	■	■	■	6 years, 5 months: 10/11–3/18	\$10
Georgia (SWOG)	Eurasia	Labor organization capacity	Training and technical / legal capacity-building	■	■	■	4 years: 12/14–12/18	\$0.9
Georgia (ICLLG)	Eurasia	Inspection and enforcement capacity	Technical support and training for labor-law enforcement	■	■	■	6 years, 3 months: 12/31/2013–3/31/2020	\$3
Haiti (MCB)	LAC	Inspection and enforcement capacity	Technical support and training for inspection, mediation	■	■	■	3 years, 9 months: 3/1/2014–12/31/2017	\$2.4
Haiti (BWH)	LAC	Labor-law compliance	Advisory services, training, compliance assessments	■	□	□	10 years, 10 months: 6/1/2009–4/1/2020	\$13.3
Jordan (BWJ)	MENA	Labor-law compliance	Advisory services, training, compliance assessments	■	□	■	3 years: 7/1/2014–6/30/2017	\$4.13
Malaysia (LL-IRR)	SE Asia	Labor-law reform and inspection	Technical assistance to reform labor laws, trainings	■	■	■	3 years, 6 months: 10/1/2016–3/30/2020	\$1.6
Mexico (IGUALDAD)	LAC	Gender equality	Technical assistance, training, and outreach	□	□	■	3 years, 4 months: 12/2014–4/2018	\$1.4

Country and project acronym	Region	Focus	Main activities	Primary targets ¹			Length, start and end dates	Budget (millions)
				Employers	Workers	Government		
Morocco (Wad3éyati)	MENA	Gender equality	Technical assistance, worker / community supports	■	□	□	4 years, 3 months: 12/31/2013–3/31/2018	\$1.3
Peru (PLIP)	LAC	Labor-law compliance	Technical assistance and training			■	4 years, 6 months: 12/2014–6/2019	\$2
Peru (BUCCPEP)	LAC	Protecting workers' rights	Technical assistance and training		■		2 years: 12/2015–12/2017	\$1
Philippines (BC-PLIP)	SE Asia	Labor-law compliance	Technical assistance and training	□	□	■	4 years, 8 months 12/2014– 8/2019	\$1.3
Vietnam (ILO-IR Phase II)	SE Asia	Labor-law reform and compliance	Technical assistance and training	■	■	■	3 years, 11 months 10/2012–9/2016	\$3.74
Vietnam (BWV-UCD)	SE Asia	Labor-law compliance, industrial relations and union capacity	Training, advisory services, dialogue facilitation	■	■		6 years: 11/2011–11/2017	\$2.3

¹ The solid squares (■) indicate the primary targets of the project. Where projects engaged other partners (but to a lesser degree), the table shows hollow squares (□). Blank cells indicate insubstantial or no engagement with that partner. Tripartite projects are represented with three solid squares if tripartitism is a core strategy and projects engage meaningfully with all three stakeholder groups. Tripartite projects may also be represented with a combination of solid and hollow squares if they do not engage meaningfully with all three stakeholder groups. The variation seen in this table represents the variation in the ways in which projects may use a tripartite strategy.

² This amount complements \$3.6 million provided by the governments of Canada, Netherlands, United Kingdom, Switzerland, and France.

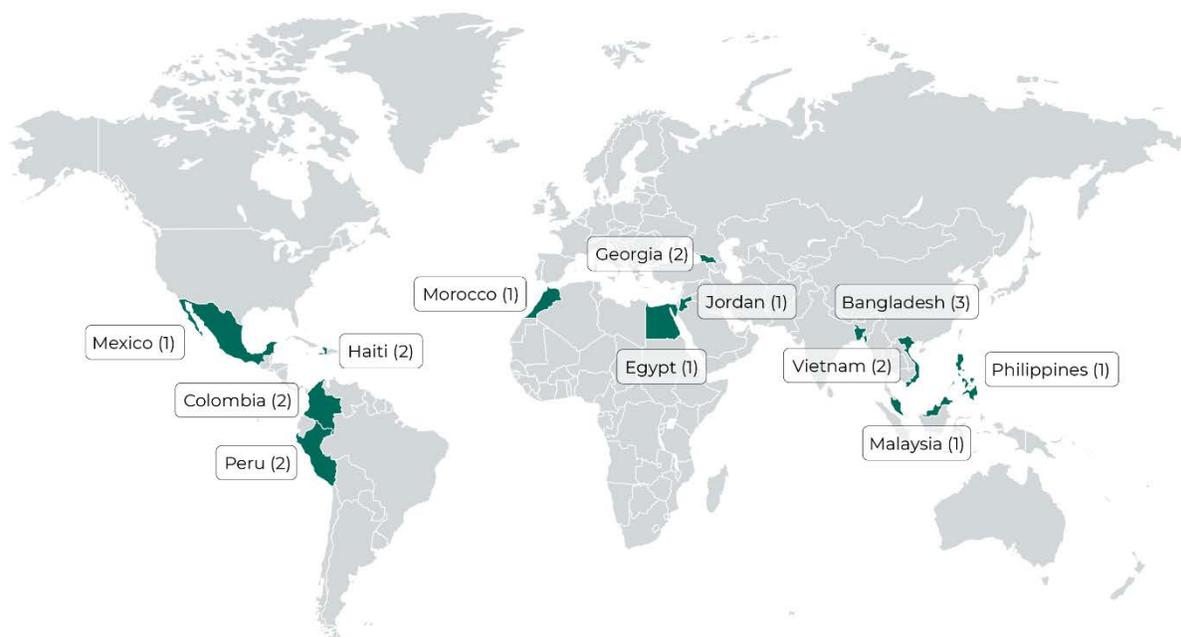
³ USAID funded BWJ from 2009 to 2014.

⁴ This amount complements \$669,789 provided by ILO and USDOL Better Work Program.

1. Project locations

Projects included in the report represent most regions of the world outside Sub-Saharan Africa and high-income countries. The 19 projects reviewed for this evaluation focus on improving compliance with and enforcement of national labor laws and international labor standards in 12 countries across Latin America and the Caribbean (7 projects), Southeast Asia (4), Middle East and North Africa (3), South Asia (3), and Eurasia (2). Countries with multiple projects reviewed include Bangladesh (3), Colombia (2), Georgia (2), Haiti (2), Peru (2), and Vietnam (2). Egypt, Jordan, Malaysia, Mexico, Morocco, and the Philippines each had 1 project covered in this review. Figure III.1 shows the geographic distribution of the projects included in this review.

Figure III.1. Number of projects included in the synthesis review, by country



Note: Figures in parentheses indicate the number of projects reviewed in each country.

2. Target stakeholders

Projects engaged with government, employer, and worker stakeholders to achieve diverse objectives. Within each country, projects that engaged government partners tended to do so at the national level by working with ministries of labor, inspectorates, or judicial systems, with minor regional government engagement. Projects targeting workers did so through national-level unions or through individual worksites (such as factories, mines, or agro-export facilities) and several projects engaged with local or regional union contacts as well. Projects working with business partners did so through national employer associations or local outreach to individual worksites. Buyer associations (key partners in the four Better Work projects examined in this review) draw together international purchasers to support demand-side compliance incentives. Several projects engaged workers and communities directly through mobile outreach as well.

Six projects emphasized an explicit tripartite approach (working with government, labor, and business sectors) and 3 projects followed through with a balanced tripartite approach. Five more projects used a limited non-explicit tripartite approach, meaning 14 of the projects practiced some degree of tripartism through their engagement with all three stakeholder areas.

3. Funding and period of performance

Projects funding and periods of performance varied, with most lasting three to five years with budgets of \$1 to \$4 million. The cooperative agreements awarded for projects reviewed in this report range in value from \$0.9 to \$13.3 million, with an average value of \$3.6 million. Allocating each project's budget evenly across its period of performance, funding amounts varied similarly from \$0.22 to \$2.09 million per year, with an average funding level of \$0.75 million per year (see Figure III.2). Periods of performance ranged from 2 to 10.8 years, averaging 4.6 years in length (see Figure III.3). Eight projects

received a total of 10 no-cost extensions, and 5 projects received one funded extension each. Finally, 7 projects evaluated and reviewed for this report were single phases of larger, longer-term projects which had preceding or succeeding phases.

Figure III.2. Project funding

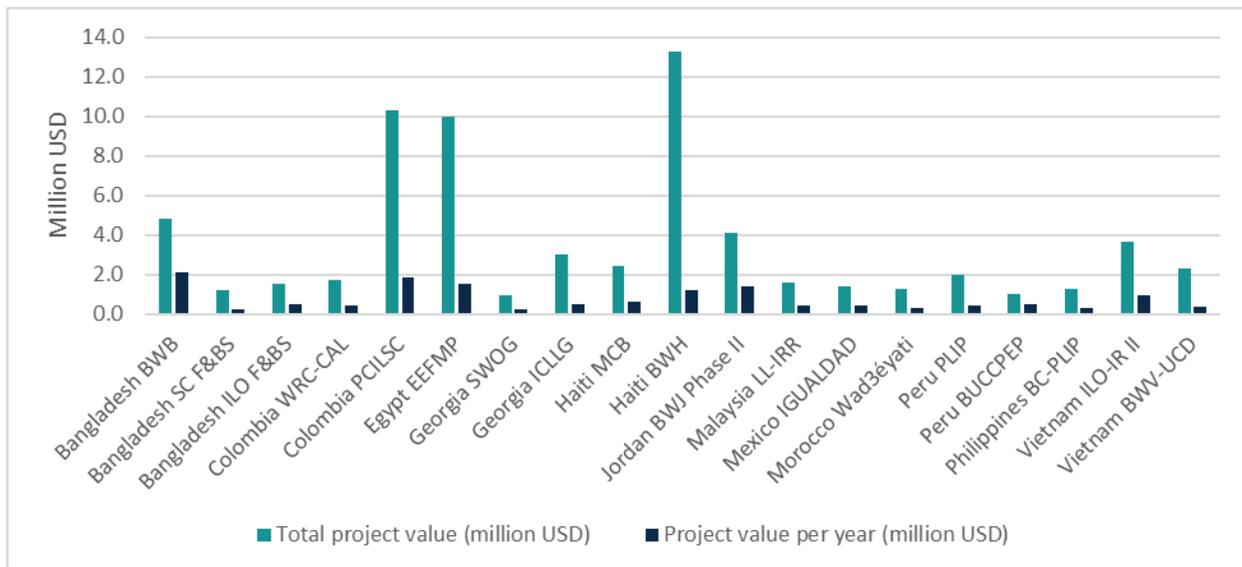
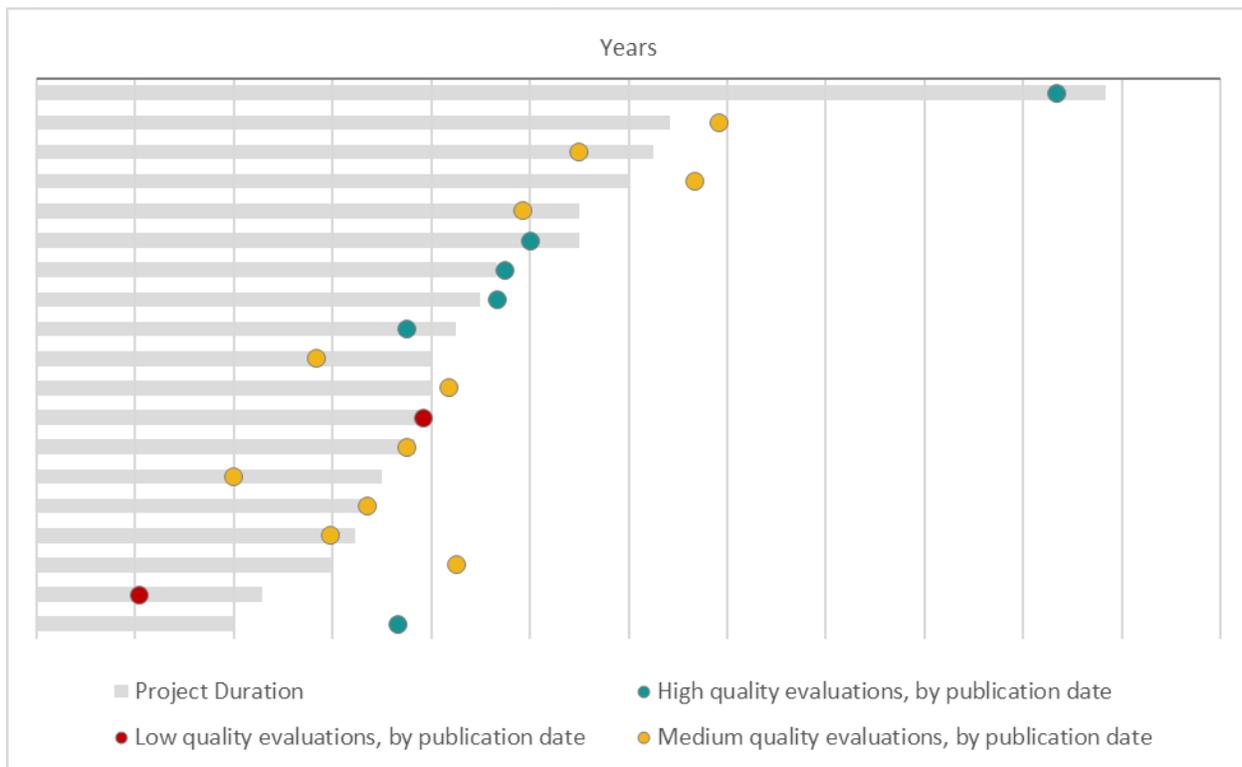


Figure III.3 Project periods of performance (years) and evaluation publication dates⁷



⁷ BWB was preceded by a design phase and BW preparation phase, both funded by DOL.

4. Project objectives and target industries

Nearly all projects pursued more than one ultimate (or long-term) objective, with the development of union capacity, government capacity, and employer compliance the most commonly observed objectives. Fifteen⁸ of the projects sought to develop union capacity or improve protections of worker rights or worker safety; 11 projects aimed to reform labor law or policy or build capacity of government agencies for inspection or enforcement of labor laws; and 11 aimed to improve employer compliance, competitiveness, or industrial relations. Better Work projects, based on a set of activities formulated by the International Labour Organization (ILO), focused principally on the first and third set of objectives.

Apart from those main objective groups, two projects sought to reduce gender discrimination and improve gender equality in the workplace (see Box III.3). One of the projects reviewed for the report targeted migrant labor issues by engaging an industry with 75 percent of the labor force composed of migrant workers. Two projects sought to reduce the vulnerability of illegally contracted workers, one project had the unintended positive consequence of supporting enforcement of laws against child labor, and no projects explicitly targeted afro-descendent or indigenous labor groups. See Box III.4 for a full discussion of projects targeting vulnerable groups.

The industry targeted most commonly (across 7 projects) was the ready-made garment (RMG) sector. RMG factories are important employers and drivers of economic growth in Bangladesh, Jordan, Haiti, and Vietnam, and projects in those countries (particularly the Better Work projects) tended to focus on improving labor law compliance in the RMG sector. Other targeted sectors included textiles, food processing, mining, public sector, ports, and agro-export sectors such as sugar cane, flowers, and palm oil. Seven projects did not have focus industries.

5. Principal inputs and activities

The main inputs and activities of the projects reviewed for this report were training and technical assistance. These components also generally comprised the bulk of each project's budget. Other activities included launching data management platforms, harmonizing regulations and reforming policy, and developing industrial relations. The following paragraphs detail those inputs and activities.

Training: All projects reviewed for this report included training inputs and activities.

- For projects aiming to **build government capacity**, training of inspectors, judges, and prosecutors generally focused on inspection procedures, reporting systems including electronic case management systems (ECMS), national labor laws and international labor standards, sanctioning tools, and prosecution practices. Inputs related to these trainings generally involved curricula, inspection checklists, and guidebooks for participants.
- For projects designed to **directly support employers with labor law compliance**, training of factory managers and business association leaders dealt with such topics as labor laws and standards, compliance requirements, assessments, safety and remediation, and industrial relations and social dialogue. Inputs related to these trainings included developing curricula and assessment tools, advisory services, and recruiting factories.
- Finally, projects focused on **worker and organized labor capacity-building** generally trained union leaders and workers on topics such as labor laws, union governance, workers' rights, life skills,

⁸ Projects with more than one objective appear in more than one of these groupings.

collective bargaining, safety standards, processes for filing complaints and seeking remediation, and industrial relations and social dialogue. Inputs for labor trainings were curricula and recruitment of union officials and of workers through either unions or factories for participation in project trainings.

Technical Assistance: All projects reviewed for this report included some version of technical assistance or advisory services.

- For **government** stakeholders, projects mainly provided technical assistance by co-developing inspection tools and checklists with ministries of labor and labor inspectorates.
- For **employers**, technical assistance included supporting factories by conducting assessments of labor law compliance, providing compliance assistance, training and advisory services, and facilitating bipartite enterprise level worker-management committees (especially through Better Work projects)
- For projects targeting **workers and union leaders**, technical assistance included direct support and accompaniment to workers in filing complaints of rights violations, seeking remedy and mentorship for labor leaders on advocacy and union growth and governance, and offering guidance on collective bargaining, conflict resolution, and social dialogue.

Launching data platforms: 11 of the 19 projects involved developing, upgrading, adapting, or launching data management or reporting platforms, including electronic case management systems (ECMS) (see Box III.1). For government partners, these were typically inspection reporting platforms. For employers, these data platforms included assessment databases where factories and buyer associations could access past reports. For workers and unions, these platforms were used to track and monitor rights violations.

Box III.1. Electronic case management systems (ECMS)

Among projects that targeted ministries of labor, most sought to develop, upgrade, adapt, or launch electronic case management systems (ECMS). ECMS are typically designed to provide inspectors and other government personnel with (a) a portal to input data (such as inspection notes), (b) a means to follow up on documented labor violations, and (c) an easy-to-use system for aggregating, processing, and reporting information on labor compliance. The following projects offer examples of successes and challenges in ECMS activities.

- **Colombia PCILSC:** this project engaged the MOL Labor Inspectorate in developing and testing an electronic Labor Inspection Case Management System with a goal of improving “the effectiveness, efficiency and transparency of the labor inspection process,”—particularly by offering capacity for “the registration and analysis of labor inspection data and reports and other services.”
 - Inspectors trained on the system were satisfied, mostly, but some were not comfortable with the use of information and communications technologies more generally. The full release of the ECMS was delayed, and the evaluation recommended that the project (in its final six-month extension) prioritize the rollout, adoption, and technical troubleshooting of the system in close partnership with the MOL.
 - MOL leaders still identified the system as the most important contribution of the project.
- **Peru PLIP:** this project sought to redesign the Labor Inspection Information System (*Sistema Informático de la Inspección de Trabajo*, SIIT) “to increase the effectiveness and efficiency of inspections.”
 - The existing system “was incomplete, not regularly updated, and lacked specific information on the remedial process around cited violations.” The National Superintendence of Labor Inspection (*Superintendencia Nacional de Fiscalización Laboral*, SUNAFIL) lacked IT-capable personnel and software to effectively strengthen the SIIT.
 - Another project objective was to have SIAN (*Sistema de Información Articulado Nacional*, National Articulated Information System) integrate external (i.e. enterprise or sector data, from employer-provided *planillas electrónicas*) into the inspector-driven system of SIIT. This was not feasible given the project and target government resource limitations.
- **Philippines BC-PLIP:** The project strove to improve utilization of the inspectorate’s management information system (MIS), originally created by the ILO under a U.S. State Department grant.

- Under this project, implementers used business process reviews to strengthen the MIS, while technical improvements and dedicated resources allowed DOLE to bring the system in-house and develop its capacity. It also allowed regional offices to access regional information.
- Inspectors trained on the new system were generally satisfied with MIS and technical support related to it. According to DOLE leadership, the MIS informs programming and policy with information on “industries with low compliance rates, regularization data, and OSH.” Other agencies also now use MIS data.

Harmonizing regulations or reforming policy: 5 of the 19 projects aimed to harmonize divergent labor regulations or reform them to align with international labor standards (ILS), usually through expert support to government agencies or union leadership in proposing legislation or policy changes.

Fostering industrial relations and dialogue: 5 of the 19 projects sought to improve industrial relations with trainings for employers or workers, development of social dialogue mechanisms or organizations, or space for workers and employers to communicate and cooperate to identify and resolve issues, such as factory-level improvement committees.

6. Implementers

The ILO or Solidarity Center implemented 15 of the 19 projects. The ILO was the primary implementer of 12 of the 19 projects reviewed for this synthesis and brings global reach and expertise in international labor standards and tripartism. Solidarity Center implemented 3 projects, and the National Union School (ENS, local to Colombia), Management Systems International (MSI), Heartland Alliance International (HAI), and Social and Human Capital (CHS, local to Peru) in partnership with Labor Development Program (PLADES, local to Peru) each implemented one project. In several projects, ILO worked with subrecipients to implement the projects.

B. Evaluation reports and data available for the synthesis review

Section summary: Evaluation reports and data

- Evaluations reviewed for this synthesis were medium-quality and varied substantially in level of detail
- All the evaluations reviewed were mixed methods, non-causal performance evaluations
- Evaluations relied on qualitative data collected through document reviews, key informant interviews, or focus group discussions and on quantitative data collected through project performance monitoring plans (PMPs) and partner data records
 - Qualitative data were subject to several forms of bias
 - Quantitative data provided through PMPs were not consistently reported
 - Validity of final evaluation findings was threatened by the timing of the evaluations (e.g. before projects concluded, without a final retrospective view of the completed projects)
- Evaluations often lacked discussion of gender issues, project budget and efficiency, and resource allocation of target stakeholders

The 19 USDOL-funded projects selected for this synthesis review were assessed in independent evaluations managed by USDOL (13 projects) or the ILO (5 projects). Key evaluation characteristics are described below.

- **The projects were largely evaluated by independent firms headquartered in the U.S.** Of the 19 projects, 6 were evaluated by O’Brien and Associates International (based in the U.S.), 5 were evaluated by IMPAQ (U.S.), 1 was evaluated by Management Systems International (U.S.), 1 was

evaluated by Universal Management Group (Canada), 1 was evaluated by *Sistemas, Familias, y Sociedad* (Peru) and the remaining 5 were evaluated by independent consultants.

- **Most project evaluations were published near projects' close.** One evaluation was published halfway through the evaluated project phase, 7 evaluations were published near the project's end (within a year), 7 were published at or close to the project end, and 4 were published after project close.
- **Evaluations often ran on a short timetable.** Within the space of 2-5 months, evaluators conducted a desk review of project documents, engaged in 1 to 2 weeks of field data collection, analyzed information, and compiled findings into a report.
- **Evaluation products included reports and infographics.** Evaluation reports averaged 80 pages, including 25 pages of appendices, and generally covered the project context and motivation, the theory of change, fidelity of implementation, and effectiveness of the intervention.
 - **Some evaluations lacked these key elements.** Several reports lacked clear descriptions of the project's results frameworks, theory of change, assumptions, challenges, or solutions.
 - **Some evaluations were supplemented with infographics.** These short documents conveyed project details and high-level evaluation findings in visually appealing formats.⁹
- **Evaluations drew data from desk reviews, key informant interviews, focus group discussions, rapid scorecards and small surveys, observations, project performance monitoring plans (PMPs), and administrative data from partners.**

The following sections review the evaluations' quality, the methods used by the evaluators, the corresponding performance monitoring plans (PMPs), gender-specific data availability, and other data issues.

1. General evaluation quality

Most of the 19 evaluations reviewed for this synthesis were of medium quality. Based on an assessment of completeness, clarity, and validity¹⁰, 2 evaluations were deemed low quality, 11 were medium quality, and 6 were high quality. This variation is important to note, because the quality of an evaluation may influence the portrayal of the project. In fact, this review found that lower quality evaluations were more likely to report that project inputs, outputs, and outcomes were achieved. It is possible that lower-quality evaluations were more likely to overstate their findings, but this review did not have access to additional sources to assess that possibility.

Evaluation detail also varied substantially. Multi-project evaluations offered less detail than evaluations focused on a single project. While these multi-project evaluations may have provided evaluators with a more holistic view of USDOL interventions in country, in Bangladesh, Malaysia, and Vietnam, the combined approach also limited the evaluators' attention to the project in question.

⁹ Since 2018, OTLA has required evaluators to produce infographics accompanying their reports.

¹⁰ Evaluation completeness, clarity, and validity were scored using ordinal systems (e.g. from 0-2) based on our review of the report. For example, evaluations with poor validity, such as those without representative samples of interviewees and focus group respondents, received a data validity score of 0 or 1. Those evaluations with stronger validity, such as those with more representative samples and stronger validation practices, received data validity scores of 2. This system does not provide a comprehensive review of the quality of evidence, but rather shows the quality of key evaluation characteristics that may influence findings.

2. Typical evaluation approach

Evaluators conducted mixed methods analysis to describe projects' implementation approach and outcomes, challenges faced, and solutions used. None of the evaluations included in this synthesis used analysis methods that allow for the identification of projects' causal impacts on outcomes of interest. This was due to contractual and budgetary limitations faced by the evaluators, who were selected and began their work after projects had already begun when it was too late to randomly assign the interventions and not feasible to create a randomized control group. With those limitations, evaluators (and hence the authors of this synthesis review) cannot confidently infer which factors caused the outcomes observed. Nonetheless, evaluators combined qualitative data with quantitative program data from the PMPs (discussed in more detail in sub-section 4 below) prepared by the implementers to explore plausible relationships between inputs and ultimate outcomes. In some cases, evaluators also collected data with small surveys and rapid scorecards to directly obtain quantitative data from beneficiaries and stakeholders, which served to triangulate findings from qualitative and program data.

To improve the rigor of future evaluations, USDOL could consider engaging evaluators earlier and for longer periods after the intervention has completed. There are three options for engaging evaluators.

1. The **first-best option** would be to bring on evaluators pre-implementation, so that they could offer insights on possible strategies to roll out an intervention while dividing target beneficiaries—or staggering their receipt of the program activities—so that the intervention's impact could be more easily quantified using a control group to estimate a counterfactual. For example, evaluators engaged early might suggest an intervention targeting workers uses an eligibility threshold to divide potential program beneficiaries, which in turn could be used to identify program impacts through a regression discontinuity evaluation design. Constructing a randomized control group or quasi-experimental counterfactual is not always possible or ethical, even if the evaluation is designed before implementation. However, options for using a counterfactual are even more limited when designing the evaluation after project implementation has begun.
2. The **second-best option** would be to have the ILAB monitoring and evaluation team draft an evaluation plan at project outset, which the evaluator, contracted at project close, then conducts.
3. The **third-best option** would be for ILAB to engage the evaluator at project close to design and conduct a retrospective evaluation.

Each of these options would deliver better results if the evaluator was retained long enough after the intervention concluded (6 to 12 months) to conduct a second, smaller data collection and assess the long-term effects of the project and its sustainability. Employer- and worker-focused projects should be evaluated using the first-best option whenever possible, given that it may be feasible in those cases for worksites and/or workers to be randomly assigned (before the project begins) to treatment and control groups and allow evaluators to assess the impact of project activities. The second- and third-best evaluation timing options could be suitable for government-focused projects, given that it is not feasible to create a randomized control group for inspectorate capacity-building or policy reform.

3. Qualitative data sources

Evaluators used a variety of qualitative data sources to cultivate a deeper understanding of project implementation, challenges faced, and solutions used. All evaluations included a desk review of project documents, and all evaluators conducted key informant interviews with stakeholders in the field.

Eleven of the 19 evaluations also conducted focus group discussions, principally among workers and other beneficiaries in bottom-up worker advocacy or employer-compliance projects. In most evaluations, the qualitative data were coded to themes and subsequently analyzed with content and comparative approaches.

While valuable, this qualitative data were subject to bias. For some evaluations, the rapid turnaround stipulated by the evaluation terms of reference required brief field visits, preventing evaluators from speaking with balanced groups of relevant beneficiaries and key informants. In other cases, when implementers selected informants for the evaluation rather than evaluators, evaluators questioned whether there was selection bias among respondents. In both interviews and focus group discussions, recall bias (misreporting past events) and response bias (offering responses affected by the evaluation itself) posed risks in most of the evaluations. Finally, for one project, the evaluator noted that interviews with workers took place on the factory floor in the presence of management staff, which risked biasing responses. Despite these limitations, qualitative data typically formed the core of the evaluations' assessments of project achievements.

4. Quantitative data sources

Evaluators drew quantitative data from two principal sources to examine the degree to which projects achieved planned inputs, outputs, and outcomes. These sources were implementers' PMPs, which provided data such as number of inspectors trained, and partner records, which included Ministry of Labor inspection reports. In 3 evaluations, evaluators also collected primary quantitative data through small surveys or rapid scorecards distributed to beneficiaries or stakeholders during fieldwork. These small survey and rapid scorecard tools allowed evaluators to quickly acquire information to complement interviews with the benefits of a) providing greater perceived anonymity to respondents and b) allowing easy manipulation of data and computation of summary statistics.

Performance monitoring plans (PMPs)

Each project team was required to submit a PMP at the outset of activities. PMPs described indicators and listed targets for planned inputs (such as number of training manuals printed), outputs (such as number of trainings held), outcomes (such as percent of inspectors using new protocols in the field), and ultimate outcomes (such as percent increase in successful remediation of labor violations). The goal of PMPs is to improve the accountability of a project to its goals and keep USDOL informed of progress toward targets through semi-annual, if not quarterly, reporting updates.

The quality of project PMPs was generally low. Of the 19 evaluations reviewed, 10 were accompanied by low-quality PMPs (with many indicators missing data, targets, or a clear description), 8 had medium quality PMPs (with some indicators missing data, targets, or a clear description), and 1 had a high quality PMP (with all or nearly all indicators adequately described and reported). PMPs varied widely in detail—some reported as few as 6 indicators, providing a sparse view of project progress, while others reported on more than 50, which burdened the evaluator and reviewer with parsing the indicators of importance. Several PMPs also offered ambiguous indicators, such as "No. of surveyed **and** % of union leaders that increase their capacity through training in negotiation, communication, dispute settlement, OSH etc. [emphasis added]", with only one value provided. Finally, evaluators for most projects stated that either poor target-setting, the absence of a project monitoring and evaluation (M&E) team member, or underdeveloped indicators (or all three factors) limited the capacity of the project teams to use the PMPs strategically to measure progress and maintain accountability to project goals.

In most PMPs, the targets set by the project were also lacking. Evaluators noted that most indicator targets were unreasonable in 8 PMPs. Ten PMPs had somewhat reasonable targets, and 1 had reasonable targets. Evaluators' main concerns with targets were that they were overly ambitious given the project scale and budget, insufficiently justified, and lacking in specificity. This report's close review of all PMPs confirms that these problems form a trend. The review also found that while some PMP targets were not revised to reflect significant barriers to or catalysts for activities, other targets were revised frequently (6 times, in one case) and may have been overfitted to the exact achievements of the project. Finally, the review found irregularities in some targets, particularly where quarterly indicator goals did not add up to life-of-project goals.

PMPs also concentrated target-setting and progress-tracking on inputs and outputs, and often lacked ultimate outcomes.¹¹ These ultimate outcomes, such as the percent of targeted factories with year-over-year reductions in non-compliance indicators, would allow the evaluator (and the reviewers for this synthesis) to quantitatively assess project progress toward goals placed near the crux of the results framework. However, several PMPs offered no indicators whatsoever for ultimate outcomes, while others planned to collect ultimate outcome data but did not follow through. For example, one labor inspectorate capacity building project tracked 14 inputs, 5 outputs, 12 outcomes, and 0 ultimate outcomes.¹² In general, projects would do well to narrow the number of results (for example, outputs and outcomes) toward the end of the results framework but should be sure to provide several clearly specified ultimate outcomes (and corresponding indicators) against which project achievements can be measured.¹³

Using PMP data to assess completion of targets

Missing data and inconsistent target-setting limit the utility of the PMP data. The methodology plan of this synthesis review proposed a quantitative assessment of projects' success using results to goals ratios (RGRs), but this metric would not be informative because of missing data in the PMPs and inconsistent target-setting. RGRs would be computed by dividing indicator results by their targets, producing a figure which shows the degree to which goals were met. By using such a measure, the review could examine the success of projects components at a granular level and readily compare the achievement of projects using similar inputs or activities. For example, assuming targets were carefully set, a project with an RGR of 1.1 for number of union leaders trained would have outdone a project with the same indicator and an RGR of 0.9, and such a contrast would merit a comparison. However, the PMP quality and targeting problems discussed in this section undermined the reviewers' confidence in RGRs as a metric of project success. Numerous indicator targets and results were missing across the PMPs, and many of those indicators provided with adequate information had results dramatically under or over what was targeted (with RGRs ranging from 0.0 to 17.0, or 1700 percent above the goal), skewing RGR averages and making comparison between projects infeasible.

¹¹ PMPs and evaluations varied widely in their use of results "level" terms (inputs, outputs, and so on). To review their quality, we processed each PMP into uniform spreadsheets and reassigned the indicators to the four main buckets: inputs - execution of planned activities; outputs - participation in planned activities; outcomes - application of new skills or capacities, and ultimate outcomes - achievement of major project goals.

¹² In contrast, Colombia WRC (the one project that produced a high-quality PMP and received M&E support from the evaluator through an earlier impact evaluation) showed an ideal narrowing of indicators: 15 inputs, 12 outputs, 8 outcomes, and 2 ultimate outcomes.

¹³ Evaluations would also be better equipped to measure long-term outcomes if they were supported in follow-up data collection 1 year after project completion.

To overcome this limitation, this review instead assessed scores for inputs, outputs, outcomes, and ultimate outcomes based on a comprehensive examination of the evaluations and the PMPs¹⁴, using a 0 to 3 scale. We found a positive relationship between our assessed scores and the RGRs, which suggests that the assessed scores are credible.¹⁵ We prefer to work with the assessed scores to incorporate a broad view of projects' inputs and outputs while avoiding being overly influenced by the extreme values found in some RGRs.

Indicators varied from project to project, impeding cross-project analysis. Each project's PMP had indicators designed to measure inputs, outputs, and outcomes specific to that project. For this review, we managed this challenge by assessing projects' success in achieving outcome targets, regardless of how they defined their outcome indicators. Within project groups (such as government-focused), using common indicators—such as inspectors showing improvement in ECMS skills—could allow for some degree of comparison between projects.

Gender data

Gender-focused data (both qualitative and quantitative) were lacking in the evaluation reports. Evaluations for 9 of the 19 projects did not discuss gender issues.¹⁶ Of those 10 evaluations that discussed gender, only 4 determined projects adequately mainstreamed gender considerations into programming (e.g. embedded gender into the theory of change or implemented gender-aware activities). Two of those 4 projects were explicitly focused on improving gender equality, and evaluations of those projects offered clearer views of the activities and approaches used.

PMPs likewise showed a lack of attention to collecting quantitative gender data. Only 9 of the 19 project PMPs had gender-focused indicators, either disaggregated from another indicator, such as *number of inspectors trained (women)*, or gender-specific, such as *number of women workers within project-targeted neighborhoods that are aware of their labor rights*. However, these indicators were often either not reported or reported without a target set, making a comparison of goals to achievements for this review impossible.

Other missing data

Three other areas of missing data limited the comprehensiveness of the evaluations and the capacity of this review to assess and compare project success.¹⁷

¹⁴ We did not have access to projects' original proposals or Pro Docs to examine the earliest iterations of the indicators and targets.

¹⁵ A complicating factor in comparing these two scoring systems was the following: where there were substantive revisions to planned inputs and outputs, we used original project targets in our scoring process if revised goals were overfitted to achievements. If new goals reflected necessary pivots to new priorities, we used those new targets in our scoring process. In contrast, we used the most current goals from the latest PMPs in calculating the RGRs. This makes the comparison of the two scoring systems less useful; here we share the charts to show the general relationship only.

¹⁶ OTLA has noted that of the 19 evaluations, only a handful had terms of reference requiring an examination of gender issues.

¹⁷ There were also inconsistencies between PMPs and corresponding evaluations. For most evaluations included in this review, the project's PMP was published months before or after the evaluation. Final PMPs were published later than evaluations published during ongoing projects, which meant the two documents presented data from different

Budget analysis and efficiency: 5 evaluations lacked even basic assessments of project budget and efficiency. Among the 14 that did explore these areas, several evaluations struggled to make concrete assessments due to poor project records of budget allocation to specific activities, and poorly reported indicators under those activities.

Resource allocation by target governments, unions, or employer associations: One useful way of assessing the influence of projects is reviewing how target institutions change funding for their programs during or after an intervention. For example, several projects reviewed for this report aimed to increase the capacity and priority of inspection services within ministries of labor. This change in priority could be reflected in increased allocations to inspectorates (although bureaucratic inertia or other budget priorities could influence allocation as well). However, few of the evaluations offered discussions of resource allocation changes in target institutions. In fact, 5 evaluations made no mention of programmatic changes in target institutions, and in no evaluation did the evaluators identify how (or whether) changes to institution programming implied changes to resource allocation.

Validity issues and limitations of evaluations

This review assessed validity of data for most of the evaluations as moderate; for 3 evaluations it was high. Validity refers to how well a given result measures what it is supposed to measure. Common constraints to the validity of qualitative data were risks of recall, response, and selection bias among interviewees and focus group discussants, while the validity of quantitative data was often constrained by indicators' ambiguity and incomplete PMPs. Evaluations receiving high scores for validity conducted thorough desk reviews and broad-reaching qualitative data collection, made efforts to address recall and response bias, had high-quality PMP data available to them, and reported using interactive validation workshops to review findings with stakeholders prior to publishing. The 16 evaluations without all these characteristics (those with moderate data validity) ran the risk of presenting results that did not measure the true achievements of the project.

The timing of evaluations relative to project completion may have also limited the validity of data and findings. As shown in Figure III.3, one evaluation was published at midterm, seven evaluations were published near the project's end (within a year) seven were published at or close to the project end, and four were published after project close. In some cases, projects were deliberately evaluated before project close, while in other cases, extensions granted by USDOL meant the project was ongoing at the time the final evaluation was completed. It is important to note that evaluations published after project close were more able to offer comprehensive and accurate views of project inputs, outputs, outcomes, challenges, solutions applied, and success of those solutions in mitigating challenges. This variation affected the capacity of this synthesis to provide a comprehensive view of all projects.¹⁸

points in time and the evaluations could not qualitatively explain the final indicator results presented in the PMPs. For this reason, in cases where PMPs were published more than 6 months after the corresponding evaluation, the reviewers used performance monitoring data from the evaluation itself to construct the RGRs. This implied both the quantitative and qualitative data derived from the evaluations for this review did not afford a full timeline view of the project.

¹⁸ The relative date the evaluation was published appears to be associated with effectiveness scores described in Section C of Chapter III. Projects with evaluations published at project midpoint or near-end tended to receive B and C scores (with only one A). Projects reviewed at project close (end) received generally mixed scores. Projects reviewed after the project close (later) received B and A scores (with only one C score). It is assumed that projects evaluated before completion have more unfinished activities and unrealized outcomes than completed projects, making comparison challenging.

Recommendations for evaluators and OTLA relating to projects' monitoring and evaluation data

- Engage evaluators at the program design stage to support the development of the most rigorous evaluation design possible
- Require evaluations to collect data at project end to allow for clearest view of project success and consider supporting evaluators in conducting follow-up data collection 1 year after project completion to measure longer-term outcomes
- Provide evaluators with the resources and time to conduct fieldwork that captures a full view of the project
- Encourage evaluators to supplement qualitative data from interviews and focus group discussions with quantitative data from rapid scorecard or small surveys
- Support evaluators in developing a representative sampling frame of potential interviewees and ask implementer to help with outreach (not final selection) to those respondents
- Require evaluators to use interactive, participatory validation workshops (virtually, if need be) with stakeholders to review findings and collect any clarifying information to improve evaluation accuracy
- Include in evaluator terms of reference (TORs) the requirement that evaluators present and dissect projects' results frameworks, theories of change, assumptions, challenges, and solutions in their evaluations
- Include in evaluator terms of reference (TORs) the requirement that evaluators assess projects' gender approaches and impacts, as well as any changes in stakeholder policies, programs, or resource allocation as a result of project activities
- Upgrade the required PMP reporting template to an easy-to-use electronic platform and require its use by all implementers to make data transparent, comparable, and useful for reviewers synthesizing trends across projects
- Make project budgets and financial reports available to evaluators (and add the review of such items to evaluation Terms of Reference) to ensure evaluators can assess project efficiency

Recommendations for implementers relating to projects' monitoring and evaluation data

- Use an upgraded electronic PMP reporting template, provide background and justifications for all indicator targets, and report all indicators on a quarterly basis to ensure data availability and project accountability
- Disaggregate all relevant indicators by gender to improve tracking of this important cross-cutting issue
- Ensure that each item in the results framework (each output, short-term outcome, and so on) has indicators established and targets set (see Appendix C)
- Adhere to the following indicator target-setting best practices:
 - 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
- Use outcomes-based budgeting systems, keep them updated, and have regular budgeting check-ins between USDOL and implementers for assessments of efficiency and calculations of cost per outcome in the project results framework

C. Factors associated with project effectiveness

To understand the conditions and structures that may have contributed to (or inhibited) project success, this synthesis examined the associations between more than 110 project factors and the effectiveness scores of each project.¹⁹ These factors include: 1) factors that OTLA can influence, 2) factors under implementer control, and 3) contextual factors. In addition, the analysis examined cross-cutting issues, such as how projects worked with gender issues and vulnerable groups, as they related to project

¹⁹ We define project effectiveness as the degree to which planned inputs, outputs, and outcomes were successfully achieved. For a full explanation of the analysis approach, see Appendix C.

effectiveness. This section presents the findings from this analysis, provides examples of how certain factors supported effectiveness, and offers deep-dive views of such factors as political will in target governments. Box III.2 describes the project effectiveness system we used and presents a summary of the distribution of scores assigned.

Box III.2: Project effectiveness and sustainability scoring system

To assess and compare projects’ effectiveness, we used an A-B-C scoring system. We assigned projects’ scores of A (high), B (medium), or C (low) for the degree to which they met goals for inputs, outputs, and outcomes. Projects could receive different scores in different areas (e.g. an A for inputs, an A for outputs, and a B for outcomes). We also scored projects for sustainability based the evaluators’ (and our) assessments of the likelihood that their impacts would be sustained well beyond project completion. Scores were assessed as follows.²⁰

Degree to which	Number of projects achieving A, B, and C scores in each dimension		
	A score	B score	C score
Inputs delivered	4	11	4
Outputs produced	4	12	3
Outcomes achieved ²¹	3	11	4
Sustainability supported	5	9	5

1. Factors OTLA can influence

OTLA has influence over the foundational aspects of projects, including funding levels, periods of performance, whether and how a tripartite approach is prioritized, and which institutions are targeted within stakeholder governments. This subsection examines the associations (and lack thereof) between these factors and the degree to which planned project inputs, outputs, or outcomes were successfully achieved.

Section Summary: Factors that OTLA can influence

- Project budget size (driven by project duration) is positively associated with effectiveness in achieving planned inputs, outputs, and outcomes
- **Projects primarily targeting government agencies (such as ministries of labor) had lower effectiveness scores than employer-focused, worker-focused, or fully tripartite projects**
- Only two projects set gender equality as a core objective, but evaluations and PMPs suggested seven projects achieved positive outcomes in terms of gender equality or women’s empowerment in the workplace
- Projects’ supporting factors, particularly being aligned with similar in-country efforts, related to higher effectiveness
- Better Work (BW) programs were moderately effective in achieving their goals, and the most successful inputs of those programs were 1) compliance assistance, advisory services, and training with factory managers and 2) factory-level committees with workers and factory managers

Total project funding appeared to have a positive relationship with effectiveness in achieving desired inputs, outputs and outcomes, but this appears to be driven by such projects’ longer duration. Small projects (with budgets less than \$2 million) and medium projects (\$2 to 5 million) had mixed (A, B, and

²⁰ Additional detail on the scoring system and the computations for quality of project management and partner and beneficiary buy-in scores is available in Appendix C.

²¹ One project was missing details on the degree to which outcomes were achieved, given that project activities had only just started at time of evaluation publishing.

C) scores on inputs, outputs, and outcomes, whereas the three large projects (those greater than \$5 million) had only A and B scores on inputs, outputs, and outcomes. However, there does not appear to be a relationship between project funding amounts *per year* and the degree to which inputs, outputs, or outcomes were successfully achieved. This implies that **the relationship between larger budgets and project effectiveness is likely driven by the role of longer periods of performance.**

Longer projects generally were more effective in achieving their goals. Specifically, projects lasting 4.5 years or longer had only A and B effectiveness scores, while projects shorter than 4.5 years had mostly B and C scores. As discussed above, having a longer period of performance was associated with having a larger budget, but the number of project extensions approved by OTLA did not have an apparent association with period of performance. Interventions that were phases of larger projects (such as Bangladesh BWB, Colombia WRC, or Philippines BC-PLIP) demonstrate higher-quality project management and better overall efficiency²² than projects that stood alone outside of any phased approach. These phased projects also took concrete steps to self-sustain financially (for example, through Better Work assessment revenues) and sustain impacts (through knowledge transfer to beneficiaries and partners) in ways that other projects did not. **The strengths of phased projects could derive simply from their longer overall duration, but a review of the evaluations suggests that the phasing processes themselves (closing out one phase’s budget and goals and opening another set) helps with planning and iteration of project activities.** Longer projects without that phasing may miss out on that opportunity to reflect on progress and set new targets. However, the evidence is mixed as to whether these phased projects also had better input, output, and outcome delivery than the stand-alone projects.

Whether a project targets government, worker, or employer stakeholders is associated with the project’s effectiveness in meeting its goals, with worker-focused and tripartite projects having higher effectiveness scores. Projects where government bodies were the primary target group tended to have lower (B and C) scores on input, output, and outcome achievement. Projects primarily targeting employers earned mixed scores and projects primarily targeting workers or using a fully tripartite approach earned only A and B scores.²³ Budgets varied substantially by stakeholders targeted for the intervention; projects primarily targeting employers and those using fully tripartite approaches tended to have larger budgets, and those primarily targeting government or workers had smaller budgets. The higher cost of tripartite projects and those targeting employers could be due to the resources required for a project to reach numerous factories in an employer or tripartite approach, versus the resources required to train a consolidated pool of inspectors or union leaders. Of note is that the level of buy-in displayed by target stakeholders did not appear related to which stakeholder group the project targeted. Table III.2

²² We define project efficiency as operating with minimal waste of resources, delivering inputs at or below planned costs, and keeping activities on track with the planned timeline.

²³ Another framework through which to examine this difference is *how* projects engage stakeholders to identify compliance issues and advance workers’ rights. Projects may use a top-down model that deploys **power over** other stakeholders through supports to government for enforcement of labor laws, particularly in the inspection of employers and imposition of sanctions. Projects may instead use **power with** stakeholders through supports to employers and workers with proactively pursuing compliance, mediation, dialogue, and collective bargaining. Finally, projects may use a **hybrid model** which deploys power over stakeholders and power with stakeholders. As expected, power-over projects are also the projects that engage primarily with government partners, such as ministries, whereas projects using power with stakeholders tend to accompany employers and workers in seeking compliance and rights-protection. Fully tripartite projects use the hybrid model. This analytical framework produces the same effectiveness associations as the stakeholder target framework.

shows general trends in budget, buy-in, project management quality, and effectiveness by primary target group.

In Table III.2, we show the general trends in projects’ budget, stakeholder buy-in, project management quality, and effectiveness by primary target group. Within each target group, there may be variation in those four aspects, so the labels of lower, moderate, and higher represent general trends. For example, some employer-targeting projects had very large budgets, while others had only medium-sized budgets, but in general the budgets were higher than the other target groups.

Projects that primarily targeted employers generally had higher budgets, moderate levels of buy-in and management quality, but mixed results for overall effectiveness. For projects primarily targeting employers, effectiveness was varied enough to merit a mixed assessment rather than moderate. Any project may also engage with other groups, but here we categorize projects by their primary target stakeholder and reserve the fully tripartite label for those projects that explicitly and thoroughly engaged with employers, government, and workers in a balanced approach. The table shows a key finding: government-focused projects tended to have lower management quality and lower effectiveness than projects targeted toward other stakeholders. Specifically, worker-focused projects and those that were fully tripartite in focus tended to have moderate management and usually achieved higher effectiveness in reaching project goals.

Table III.2. General trends in project factors and achievement by primary target group

Primary target	Budget	Level of buy-in	Management quality	Effectiveness
Employer	Higher	Moderate	Moderate	Mixed
Government	Moderate	Lower	Lower	Lower
Worker	Lower	Higher	Moderate	Higher
Fully tripartite	Higher	Moderate	Moderate	Higher

Within these four general target groups, two projects targeted gender equality as a core objective with poor success (see Box III.3) and three produced benefits for vulnerable populations (see Box III.4).

Projects working with government partners are hindered by lack of political will and government capacity, especially when targeting legal stakeholders (judges and/or prosecutors). As discussed in sub-section 3 of this section, projects across all models encounter challenges with government capacity and political will, and those specifically targeting government partners are more directly affected by such challenges. Among projects that targeted government inspectors, project effectiveness appears to relate to whether projects also targeted legal stakeholders. No projects targeted legal stakeholders without also targeting inspectors. The 10 projects that targeted inspectors had a mix of A, B, and C scores, but within that group, the 5 projects that targeted inspectors only had A and B scores, while those 5 that also targeted legal stakeholders had B and C scores. Four possible explanations for this difference arise:

- 1. Projects targeting legal stakeholders had to coordinate activities between ministries and judicial entities that had low levels of interagency cooperation.** For example, in the Colombia PCILSC project, an existing communications gap between the Ministry of Labor and the Office of the Prosecutor General may have hindered the project’s efforts to promote effective prosecution of anti-union violence and conduct.
- 2. Changing the practices of legal stakeholders may simply be more challenging than changing the practices of inspectors.** For example, evaluators noted the training of judges under the Haiti MCB

project led to increases in participants' knowledge, but there was no evidence that the increases in learning corresponded to improved handling of labor cases. In another example, labor mediation activities delivered under the Georgia ICLLG project were stymied in part by the lack of readiness among the five government partners it engaged (one of which was the Supreme Court).

- 3. Projects targeting both inspectors and legal stakeholders had poorer management and efficiency than most other projects, including those targeting only inspectors.** This implementer deficiency could have been coincidental, and while it was likely unrelated to targeting decisions, it may have hampered advances toward project goals.
- 4. Finally, projects targeting legal stakeholders may have aimed for outcomes that occurred beyond the project period that was evaluated.** For example, project efforts such as setting legal precedent for workers' protections by moving critical cases through the courts could be slowed down by judicial processes and appeals, meaning that the achievement of such a goal could not be captured in the evaluation published at project end.

Projects targeting government stakeholders may be less effective than projects targeting employers or workers because the assumptions underpinning the theory of change were false. Specifically, the projects assumed activities would not be impeded by political problems or government capacity, and that they could achieve substantial improvement in outcome areas and produce lasting change at the development objective level. However, political will to engage with project activities was often lacking, and the capacity of the partner government agencies was often inadequate to absorb and use the contributions of the projects. Finally, political turmoil or social unrest drew the attention of some government partners away from project engagement, delaying delivery of core intervention components. By not interrogating the assumptions behind each causal linkage in their results frameworks—and then preparing plans to mitigate potential challenges—these government-targeting projects fell short of their goals (see section III.C.2 for more detail on these challenges and Appendix C for guidance on results frameworks).

Box III.3. Better Work programs

Better Work (BW) programs reviewed for this report were moderately successful in achieving their goals.

BW is a program model of the ILO to support compliance with international labor standards (ILS) in ready-made garment (RMG) factories around the world. BW projects assess factory compliance with national labor laws and ILS, provide technical assistance and training for factory managers to address violations of labor standards, engage workers through factory-level dialogue committees and training, and publish factory-level assessments for international buyers to examine compliance in their supply chain. BW may also work with national government ministries to develop inspection capacity or with unions to develop governance and advocacy capacity. Finally, BW can support project activities by charging international buyers for access to factory compliance reports.

This review synthesized findings from evaluations of four BW projects:

- Better Work Bangladesh
- Better Work Haiti
- Better Work Jordan
- Better Work Vietnam

Evaluations of these projects found the two most successful BW activity areas were 1) compliance assistance and advisory services for factory managers, who learned about compliance and how to achieve it in their facilities, and 2) the factory-level committees with workers and managers, which supported constructive dialogue between parties and resolution of worker rights violations. The activity area with the most challenges was ministry engagement (due to low government capacity).

Box III.4. Projects with gender equality as objectives or impacts

Two projects explicitly set gender equality goals as core objectives, but seven projects reportedly had some level of positive impact on gender issues.

Projects with gender equality as core objectives

Two of the projects reviewed for this synthesis explicitly focused on gender equality as a core objective (Mexico IGUALDAD and Morocco Wad3éyati).²⁴ Neither displayed high-quality management or high project efficiency, and neither was successful in fully achieving planned inputs, outputs, and outcomes.

- In Mexico, the IGUALDAD project focused on reducing discrimination based on gender, sexual orientation, and pregnancy status (D-GSOPS) by building government capacity in both the Ministry of Labor and the Attorney General's office to enforce new labor laws and by increasing awareness of the laws among workers, employers, and social service organizations.
 - The project encountered lack of buy-in and support from the ministry counterparts, limited employer participation in the project, and a lack of understanding on the part of project stakeholders about how to report discrimination. These challenges derailed several activities and limited the project's progress toward its goals.
- In Morocco, the Wad3éyati project focused on improving gender equality in hiring, training, and promotions by supporting employers' compliance with labor laws and by assessing equality-supporting infrastructure (e.g. security, day care) in workers' communities.
 - The evaluation notes the project was constrained by an insufficient budget and a team that was understaffed for the objectives set. These limitations meant the project did not have enough time to allow for culture

²⁴ Better Work (BW) projects seek to improve employer compliance with four main areas of international labor standards (ILS), one of which is non-discrimination and covers gender discrimination issues. In their respective PMPs, BW Bangladesh, Haiti, and Jordan projects each reported only one gender-disaggregated indicator (number of women participating in project activities), while BWV-UCD did not track any. In their assessments of gender impacts, evaluators found BW Bangladesh had some positive effects, BW Haiti had substantial positive effects, and BW Jordan evaluation reported no positive effects. Evaluators for BW Vietnam did not report gender effects. Given these data gaps, it is difficult to assess how the BW focus on compliance with non-discrimination standards might produce gender outcomes. However, findings from Haiti (the most successful BW program in the gender space) suggest the project team addressed gender-specific concerns by working with factory managers and workers to raise awareness on gender discrimination and sexual harassment, enforce prohibitions against them, and improve maternity protection in most factories. For a general view of the Better Work projects examined in this synthesis, see Box III.3.

change in participating employers nor enough staff to engage communities effectively in the assessment activities.

Gender outcomes across all projects

While only two projects focus on gender equality as a core objective, gender is a cross-cutting issue that can and should be considered by funders, implementers, and evaluators across all interventions. However, neither projects nor evaluations gave sufficient attention to the issue.²⁵

Of the 19 projects assessed by evaluators, only 9 tracked gender-disaggregated indicators in their PMPs. Most of those indicators had no targets set (see section B, sub-section 4 of this chapter for more detail). Only 4 evaluations determined that projects had adequately mainstreamed gender considerations into their designs (e.g. embedded gender into the theory of change and produced gender-aware activities).

Of the 19 evaluations reviewed for this synthesis, only 10 discussed gender issues. Evaluations of 7 projects found positive impacts on gender equality or women’s empowerment. According to the evaluations, these positive outcomes were fostered by the following actions:

- Mainstreaming gender as a cross-cutting issue in project design
- Prioritizing and tracking women’s participation in trainings
- Including employer training and technical assistance on sexual harassment, gender discrimination, and maternity protections (example: BWH, Morocco Wad3éyati)
- Supporting policy reforms to enforce anti-discrimination and anti-sexual harassment standards
- Supporting firms in changing HR policies to reflect anti-discrimination, bullying, and harassment standards
- Recruiting NGOs and donor government bodies (for example, USDOL, US Consulates) in facilitating relationships for project activities to have maximum reach
- Supporting soft skills development and leadership among women union leaders

No strong pattern emerges connecting specific implementers to project effectiveness. We observed variation in project effectiveness within and between implementers.

Projects’ supporting factors that span OTLA areas of influence, factors under implementer control, and project context appear to have positive relationships with project effectiveness. Supporting factors included 1) local or international pressure or support for the project, 2) buy-in of local stakeholders and beneficiaries, 3) alignment with similar in-country efforts, and 4) implementer/donor flexibility. The more supporting factors a project had, the more likely the project was to be successful. The supporting factor that was most closely associated with effectiveness scores was alignment with other projects or ongoing efforts (such as continuing work of previous project phases). Projects with that supporting factor received principally A and B scores for effectiveness, whereas those that did not have alignment or continuity with other efforts received B and C scores. Several evaluations noted that synergies between the project at hand and other efforts, from international campaigns to local projects, allowed project activities to make a greater impact.

²⁵ Few evaluations in this pool had Terms of Reference requiring evaluators to explore questions related to gender.

Box III.5 Projects engaging vulnerable populations

Four projects sought to directly support populations in precarious or illegal labor arrangements, including migrant workers and illegally contracted workers. One project unintentionally addressed child labor problems. No projects sought to support indigenous or afro-descendent groups specifically.

- **BW Jordan was the only project among those selected for review that targeted migrant workers, and it appeared to have moderate positive effects with that group.** The project supported Jordanian RMG factories (the labor force of which is 75 percent migrant workers from South Asia, and mostly female) with compliance assessments and technical assistance. The project was moderately successful in delivering planned inputs and outputs (such as assessments performed and union staff trained on dialogue), and less successful in producing the core desired outcome, employer compliance. The project did help produce a unified contract to protect migrant workers' rights.
- **Evaluators of the Peru BUCCPEP and Peru PLIP projects noted employers often use illegal subcontracts in hiring, and both are working to address that issue.** Illegally contracted workers are vulnerable to arbitrary terminations, abuse, and intimidation, among other threats. The BUCCPEP project supported union federations and the labor inspectorate in advancing a court case which could set a precedent preventing employers from using such contracts. The PLIP project sought to improve the capacity of labor inspectorate in identifying and reporting illegal and abusive contracts.
- **Evaluators of the Philippines BC-PLIP project found that a new data system could help address child labor** (though that was not its original purpose). Specifically, the management information system (MIS) developed by the project improves transparency and accountability throughout the government inspection system and is now used to alert the inspectorate on child labor law violations.

Recommendations relating to factors under OTLA influence

- Consider funding longer-term interventions with appropriately sized budgets to achieve maximum progress toward project goals
- Prioritize flexibility with implementers as they combine, expand, narrow, or realign core project activities to address evolving risks and persistent challenges
- Consider increasing the share of OTLA projects where the primary target group is workers, unions, or their federations, given these projects appear to be relatively low in cost and relatively high in effectiveness
- Consider developing a ready toolbox of strategies to mitigate challenges with political will and government capacity in projects that target government actors (see sub-section 3 in this section)
- Consider requiring projects to mainstream gender considerations into their designs to improve gender equality outcomes
- Consider investing in projects (and advocating for project activities) that align with similar efforts in the target country to produce synergies and expand impact

2. Factors under implementer control

In project planning and implementation, implementers have influence over design and activities, management structure, efficiency, risk mitigation, and sustainability planning. This subsection examines the associations (and lack thereof) between these factors and the degree to which planned project inputs, outputs, or outcomes were successfully achieved.

Section summary: Factors under implementer control

- Projects with high-quality management and high efficiency showed greater effectiveness
- Projects that anticipated risks but did not address them had lower effectiveness
- Project design strengths such as donor and implementer flexibility supported effectiveness, while design weaknesses such as susceptibility to delays caused by external factors were connected to lower effectiveness
- Explicitly tripartite and explicitly non-tripartite projects were more successful than projects that made limited commitments to a tripartite approach
- Projects with the most coherent, evidence-based theories of change were the most effective

Management, efficiency, and risk

Projects with high-quality management and high efficiency showed greater effectiveness. Well-managed, efficient projects generally earned A and B scores for achieving planned inputs, outputs, and outcomes, whereas poorly managed and less efficient projects achieved a higher proportion of C scores. As noted earlier in Section C of Chapter III, projects with longer periods of performance also had a higher proportion of A and B scores than projects with shorter durations. Projects that are well-managed may be more likely to have longer periods of performance for three reasons. First, well-managed projects may be more likely to be extended and less likely to be discontinued. Second, OTLA may have selected implementers known to be good managers specifically for projects with longer durations. Third, projects with longer durations may simply allow for implementers to refine management structure and streamline their activities.

Projects with lower quality project management and efficiency were more susceptible to external factor delays (from political turmoil, for example), but some projects and OTLA used effective strategies to handle such delays. We are unable to say whether projects' poor management makes them more vulnerable to external delays or whether external delays make projects more difficult to manage. Well-managed projects, such as Colombia WRC, minimized their susceptibility to external factor delays by flexibly selecting partners (for example, regional and local union bodies instead of national ones) to aid in outreach to beneficiaries. Other projects structured their work in such a way that allowed them to be flexible in the face of external factor delays, avoiding derailing activities. For example, Egypt EEFMP faced political uncertainty in the form of high government official turnover and low trust between partners. OTLA acknowledged the difficulties and remained flexible with the implementer, who in turn strengthened communication with partners to build trust and project momentum.²⁶

Projects that did not address risks (as identified by evaluators and reviewers) had lower effectiveness. Specifically, the more risks a project faced but did not mitigate—such as contractual impediments²⁷ or the potential backlash against workers for participating in the project—the less likely the project was to achieve its planned inputs, outputs, and outcomes. Across the 19 projects:

- Lack of political will in government partners was a common risk and often unmitigated. Six of the eight projects facing this risk did not adequately mitigate it. Lack of employer will and lack of worker, union, or community will were much less common and more easily mitigated.
- Political turmoil or sensitivity (of the project issues) and the lack of government capacity were also common risks and were more likely to be anticipated and not mitigated, or unanticipated and unmitigated, than anticipated and mitigated. Specifically, 6 of the 9 projects facing political turmoil or sensitivity did not mitigate the risks, and 8 of the 10 projects facing low government capacity did not adequately mitigate that risk.
- Poor trust among project stakeholders, or low understanding of project priorities, went unmitigated about half of the time the challenge manifested. The same is true of contractual challenges, which included delayed release of project funds, ILO financial reporting obligations, conditions placed on

²⁶ See section 4.4 for more detail on common challenges and how projects dealt with them.

²⁷ Contractual impediments included delayed release of project funds, ILO financial reporting obligations, conditions placed on travel, difficulties coordinating efforts with subcontractors, and poor communication among contracting partners.

travel, difficulties coordinating efforts with subcontractors, and poor communication among contracting partners.

- Inadequate union capacity was difficult for relevant projects to mitigate, and natural disasters and low project uptake among individual beneficiaries, while rare, were unmitigated.

Projects with high levels of unmitigated risks tended to have deficiencies and design weaknesses such as inadequate funding or project reach for the targeted problem, susceptibility to external factor delays, and poor project management and accountability systems.

Project implementers appeared to anticipate risks more often than they were caught by surprise.

Commonly anticipated risks included political turmoil and lack of government capacity, whereas commonly unanticipated risks included lack of union capacity and lack of political will. However, most evaluators did not explicitly state whether implementers had made adequate efforts to understand project contexts during the design phase and had subsequently adjusted their activities during implementation.²⁸ It is also worth noting that a project anticipating a risk did not imply the risk would be successfully mitigated.

Design strengths, weaknesses, and approaches

Five design strengths each appeared associated with project effectiveness and/or project management and efficiency. Taken together, the total number of design strengths per project was also positively related to the project's success. The design strengths associated with good management, efficiency, and effectiveness scores were the following.

1. Donor and implementer was flexible with adjustments to activities
2. Project structure built in mechanisms for impacts to self-sustain
3. Project was aligned with other projects in-country²⁹
4. Project planned to transfer expertise and activities for local ownership
5. The project explicitly employed a tripartite approach

Two other design strengths (having a training approach carefully calibrated to local needs and capitalizing on existing will of stakeholders to expand activities) did not have apparent relationships with effectiveness.

Two project design weaknesses were each associated with lower effectiveness in reaching planned goals. As with strengths, the total number of design weaknesses per project was also negatively related to the project's success. The weaknesses associated with effectiveness were 1) susceptibility to external-factor delays (such as approval processes from ministries of labor) and 2) poor project management and accountability systems. **Susceptibility to external-factor delays, in turn, was associated with projects being top-down, enforcement models and with the challenges of low political will, high turmoil and low government capacity.** Having a poor project management and accountability system was also more common in projects with smaller budgets and shorter durations. Implementers with low capacity

²⁸ DOL and implementers may wish to refer to Appendix C for guidance on interrogating assumptions during the project design phase.

²⁹ This design strength in turn appeared to be related to country income; projects aligned with similar efforts took place in countries with median GDP per capita less than half that of countries where projects took place without alignment to other projects.

(understaffing, a lack of expertise, or a lack of on-the-ground connections) also tended to yield low output and outcome scores.

Projects’ use of a tripartite approach was associated with effectiveness. As noted above and in the previous subsection on factors that OTLA influences, projects varied in the degree to which they engaged government, business, and labor groups. A total of 14 total projects either explicitly prioritized a tripartite approach or simply engaged all three groups and thus were *de facto* tripartite. Overall, these 14 projects had lower effectiveness scores than the 5 non-tripartite projects. However, the relationship between a tripartite approach and project effectiveness appears to be driven by *how explicitly* the approach is emphasized and pursued. Projects that worked with all three partner groups but *did not emphasize* the approach as a core strategy (were *de facto* tripartite) had lower effectiveness scores than both *explicitly tripartite* and *non-tripartite* projects. Specifically, the review found that:

- The 5 non-tripartite projects had only medium and high effectiveness scores (A and B), while
- The 6 explicitly tripartite projects had principally medium and high (A and B) scores, followed by low (C) scores, and
- The 8 non-explicit, de facto tripartite projects (6 of which were ILO-implemented) had nearly all medium and low (B and C) scores.

A key factor associated with project success (as discussed in the previous subsection on factors that OTLA influences) and with tripartitism was budget size—explicitly tripartite projects tended to have larger budgets than non-explicit, de facto tripartite projects. This could suggest that 1) resource availability determined the degree of a project’s tripartite emphasis, 2) that the initial project design informed the award amount, or that both factors influenced each other.

Theories of change, logic models, sustainability, and effectiveness

ILAB requires projects to express theories of change in results frameworks (RFs).³⁰ To explore the strength of each theory of change, we examined whether the quality of a project’s theory of change was related to its success in meeting its goals. Our analysis found that projects with the strongest evidence-based, theory-driven approaches had only A and B effectiveness scores, whereas projects with theories of change that were poorly formulated had mixed scores.³¹ We also examined whether the specific inputs, outputs, immediate outcomes, and ultimate outcomes that projects planned as part of their logic model (such as *develop data reporting system* [a common input] or *government inspectors trained* [a common output]) had relationships with the effectiveness of projects in meeting those goals. We found few strong associations. Note that the findings described below refer to projects’ *planned* inputs, outputs, immediate outcomes, and ultimate outcomes rather than observed results.

Projects’ planned inputs which we grouped broadly into 1) development and delivery of training, 2) technical assistance offered, 3) data systems, and 4) industrial relations activities. Of these, only one input

³⁰ We were not provided with the original RFs drafted by project teams, and evaluators did not give consistent attention to RFs. Among the 19 projects, evaluators indicated RFs for 9 projects were inadequate, one project RF was assessed as adequate, 3 RFs were mentioned with no assessment, and evaluators did not mention RFs at all for 6 projects.

³¹ 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) according to evaluators’ assessments. We provide detailed guidance on formulating theories of change using results frameworks in Appendix C.

class showed an association with effectiveness: **projects that planned industrial relations inputs, such as bipartite factory committees and social dialogue centers, had higher effectiveness scores than project that did not have those inputs.**

Projects' planned outputs which we grouped into 1) technical assistance received by intended participants, 2) employers or factory managers trained, 3) workers or union representatives trained or guided, 4) government officials trained, 5) data systems upgraded, and 6) buyers engaged. Of these, only two output classes showed some association with effectiveness: the group of projects that offered employer or factory training appeared to have lower average effectiveness scores than those which offered worker or union training or guidance.

Projects' planned immediate outcomes which we grouped into 1) more efficient inspections or reformed regulations, 2) data tracking systems in use, 3) factories or buyers working on compliance or dispute resolution, 4) workers or labor leaders knowledgeable on rights and advocacy, and 5) women entering the workforce. **None of the immediate outcomes had apparent relationships with overall project effectiveness.**

Projects' planned ultimate outcomes which we grouped into 1) improved government capacity for inspection and enforcement of labor laws and worker protection, 2) unions, workers, and social partners more active in advocacy, 3) employers more knowledgeable with higher compliance, 4) long-term financial, institutional, and political viability of the program, 5) improved gender equality, and 6) improved dialogue, remediation, and industrial relations. Among these ultimate outcomes, projects seeking greater union, worker, or social partner activity were more likely than other projects to have high effectiveness scores, while projects seeking long-term program viability appear to have fewer A scores for effectiveness than other projects.

Evaluators' assessments of sustainability did not appear related to projects' effectiveness scores. Drawing from evaluators' discussions of projects' plans for sustainability, we scored each project with an A if financial and institutional plans were laid to ensure project components or impacts were mostly sustained, a B if some project components or impacts were likely to be sustained, or a C if few or none of the project components or impacts were likely to be sustained. There were five projects that received A scores, nine that received B scores, and five that received C scores. Those projects with A sustainability scores tended to have revenue-generating activities that would continue beyond the project (such as Better Work assessment subscriptions) and strong government buy-in and ownership of future activities, follow-on funding through other bilateral donors or multilateral organizations, or strong labor federation commitments to maintaining capacity built through the project. A key finding was that sustainability scores did not appear to be strongly related to whether the project had effectively delivered planned inputs, produced planned outputs, or achieved planned outcomes. Indeed, some of the projects that received A effectiveness scores received C sustainability scores, and evaluators noted in such cases that implementers paid insufficient attention to the continuity of impact beyond the project period of performance. However, nine of the evaluations reviewed for this synthesis were published prior to project completion, which could have undermined the accuracy of evaluators' assessment of the sustainability of project outcomes.

Beyond the inputs, outputs, and outcomes described above, projects with higher levels of effectiveness tended to use the following nine strategies and activities:

1. Building inter-party structures, such as bipartite Performance Improvement Consultative Committees (PICCS) under the BW model or tripartite committees on wages and policy

2. Building union leadership capacity and training trainers within unions
3. Providing direct supports to workers for self-advocacy and legal action.
4. Supporting workers in filing claims and pursuing cases that can set precedent for protections of workers' rights.
5. Strengthening the regulatory framework of labor law enforcement, developing intuitive training on inspection tools, and supporting data systems development
6. Training inspectors for specialist expertise (specifically, best practices for assessing, monitoring and reporting on labor conditions) and gender equality issues
7. Systematically following up on trainings to maintain trainee knowledge and interest
8. Aligning project activities with international trade status goals (such as GSP) or US aid legislation to build momentum³²
9. Aligning export companies' and international buyers' incentives with workers' rights through assessments of working conditions or by setting priorities on such areas as gender equality.

Recommendations relating to factors under implementer control

- Recommendations for implementers:
 - Design for flexibility in combining, expanding, or realigning core project activities to address evolving risks and persistent challenges
 - Reach out and offer additional support to low-capacity target governments
 - Develop and refine an evidence-based theory of change with donors and stakeholders to support effectiveness (see Appendix C for more detailed guidance)
 - Draft, follow, and regularly update a sustainability plan to increase the likelihood that project impacts are maintained
- Recommendations for OTLA to support implementers:
 - Consider selecting implementers that have demonstrated capacity to execute all project components or be prepared to provide initial and ongoing guidance and support for implementers in essential areas, such as monitoring and evaluation
 - Consider committing more funding to projects facing high risks and/or using a quick-release disbursement mechanism to deliver funds in a timely manner
 - Refer implementers to infographics on 1) common challenges and solutions 2) contextual challenges, and 3) developing results frameworks, and also provide them with Appendix C for guidance on interrogating assumptions during the project design phase
 - Require strong monitoring and evaluation, PMP, and accountability systems and support implementers in these efforts with written resources and mentoring
 - To develop evidence-based theories of change, dedicate necessary resources before the solicitation stage to locate evidence related to the project objectives and examine the proposed interventions against that evidence. To facilitate repeated reviews of the evidence, USDOL may wish to compile evidence of the effectiveness of various specific intervention activities that are commonly part of USDOL-funded projects in an organized or searchable document that could be accessed when conducting evidence reviews for the theory of change of any new project. This document should be updated regularly to incorporate new evidence.

³² Whether the target country had a free trade agreement (FTA) with the U.S. did not have an apparent relationship with project effectiveness.

3. Contextual factors

Factors such as economic conditions, prevailing national attitudes toward workers’ rights, and political turmoil are beyond the influence of USDOL or project implementers, and yet could affect the delivery of labor enforcement projects. However, our review suggests that among the 19 projects, such factors did not have strong or direct relationships with effectiveness.

Section summary: Contextual factors

- Economic factors, sectors of focus, general public attitudes toward workers’ rights, beneficiary reluctance, and influence from outside stakeholders did not appear to have strong relationships with project effectiveness. Government capacity and political will, climate, and instability also did not appear to have strong relationships with project effectiveness.
 - Intervening factors—such as the quality of project management or the success of solutions applied to the challenges—may have had more immediate effects on project progress than the presence of the economic, social, political, or government factors themselves.
- Three evaluations noted that the projects they assessed worked in contexts where workers were not free to organize. They addressed that issue through policy reform supports, union capacity development, and contracting guidance with moderate to high success.

The following factors did *not* appear to have strong, direct relationships with project effectiveness in achieving planned inputs, outputs, and outcomes.

- The gross domestic product (GDP) per capita, unemployment rates, and Gini indices (income inequality) in target countries as of project start date
- The sector in which each project was focused, such as RMG, agricultural processing, or general private employers with no specific sector
- The state of public attitudes toward workers’ rights (assessed as negative, neutral, or positive)
- The degree of union, employer, or community reluctance to engage with the project
- The degree of positive influence from prominent national and international stakeholders

As with political factors discussed in the following section, the contextual factors listed above may not have strong, direct relationships with project effectiveness because of the role of intervening factors. For example, a project’s management quality or the success of solutions it applies to contextual challenges may have had more immediate effects on effectiveness than the presence of the contextual factors.

4. The role of government capacity and political will, climate, and instability

Political factors and government capacity arose in nearly every evaluation as areas of concern. Specifically, evaluators cited challenges with political climate, political will, target government capacity, and political instability as threats to project success. These challenges generally did not have a consistent relationship with project effectiveness, which could be because intervening factors—such as the quality of project management or the success of solutions applied to the challenges—had more immediate effects on project progress than the presence of the political and government factors themselves. Each factor and its relationship to effectiveness is presented below.

- **Political climate.** Looking across the 19 projects, the general political climate in which a project took place (unfavorable, neutral or mixed, or favorable to project focus) had no strong apparent relationship with the degree to which outcomes are achieved. Only 2 projects had unfavorable

political climates, 14 projects had neutral or mixed political climates, and 3 projects had favorable political climates.

– **However, among projects targeting government stakeholders, those facing *political will* challenges from their partners (mitigated or unmitigated), made less progress toward goals.³³**

- **Government capacity.** Challenges with target government capacity were common across projects, but the presence of such challenges did not appear to have a strong relationship with effectiveness. Evaluations often mentioned that poor government capacity (such as high turnover of key staff or poor use of systems) impeded project efforts (such as relationship-building for policy reform), but our assessment of inputs delivered, outputs produced, and outcomes delivered did not find a relationship between low government capacity and project effectiveness.
- **Political turmoil.** Projects that operated in countries where there was political turmoil or political sensitivity of activities (and such risks were not mitigated) had lower effectiveness scores than projects that either mitigated the issues or did not face them.

5. Projects operating in settings where workers are not free to organize

In some country contexts, OTLA projects supported unions and workers where freedom of association was not yet fully protected. Project activities addressed these barriers with activities to reform policy, develop union capacity, and guide contracting practices, and received A and B scores in project effectiveness.

In Vietnam, the BWV-UCD and ILO-IR II project operated in a legal environment where all workers were technically represented by one government-sanctioned national union, the Vietnamese General Confederation of Labor (VGCL). Factory-level unions were close to management. For example, union leaders were also senior company executives, an overlap of roles that produced conflicts of interest in defending the rights of workers. These conditions fed into a pattern of spontaneous, wildcat strikes that took place without support of union leadership or carefully crafted demands.

- The ILO-IR project supported government partners in developing regulations to enforce new Labor Code and Trade Union legislation, which aligned Vietnamese labor laws with ILO declarations and conventions protecting union development and collective bargaining. This project was not effective in delivering inputs, but nonetheless achieved most of the desired outcomes.
- The BWV-UCD project supported the growth of union capacity, specifically through training of grassroots trade union (GTU) and regional trade union leaders on social dialogue and worker representation. This project was highly effective in delivering inputs and producing outputs, and moderately effective in achieving desired outcomes.

In Jordan, evaluators of the BWJ project noted that migrant workers (who compose 75 percent of the RMG labor force) experienced barriers to freedom to unionize, since by law migrant laborers were not allowed to form their own associations. Migrant workers have also experienced involuntary servitude, human trafficking, and discrimination based on race and origin which has translated into migrants not receiving the minimum wages secured for Jordanian workers.

³³ We exclude Malaysia LL-IRR from this analysis because it is a ministry-targeting project and so few of its activities were completed at the time the evaluation was published.

Project activities included forming factory-level committees to facilitate communication between workers and management, supporting grassroots trade unions (GTUs) in improving communication with migrant workers, and producing a unified contract to protect migrant workers. Overall, this project was moderately effective delivering planned inputs and producing planned outputs but was less successful in achieving its ultimate outcomes.

Recommendations related to dealing with contextual factors

The presence of contextual factors such as economic or political conditions does not appear to have a strong a direct relationship with project success—likely because some projects adopted strategies to address them. This implies projects should pay close attention to contextual factors that could delay or impede activities—especially issues with **target government capacity and political will**—and identify solutions early that could mitigate the effects of challenging conditions.

D. Challenges and solutions

During project delivery, many of the risks faced by implementers materialized into concrete challenges, including political and government-related challenges, internal project deficiencies, and union, worker, and employer-related challenges. This section provides an overview of the challenges that projects encountered and examines the success of solutions that projects applied.

Section summary: Challenges and solutions

- Evaluations reported that the 19 projects encountered challenges 73 times, and applied solutions 26 times.
- The five most common challenge areas faced by projects in achieving their goals were:
 - limited government capacity
 - political will or resistance
 - political uncertainty or social unrest
 - insufficient implementer capacity
 - low union or labor federation capacity
- The most effective solutions in overcoming those challenges involved increasing communication and flexibility, narrowing activities, and providing direct supports to low-capacity partners.

1. Description of common challenges

Projects faced challenges which can be grouped into political and government-related challenges; internal project deficiencies; union, worker, and employer-related challenges; and other challenges. Sub-section 2 presents examples of the five most common challenges and the solutions applied to solve them. Other common challenges included project team hiring difficulties and turnover (issues which, when they arose, were not addressed according to the evaluator), employer reluctance to participate (sometimes addressed), and backlash against workers for participation (always addressed). The complete list of challenges is presented in Table III.3.

Table III.3. Challenges and solutions

Challenges		Times occurred	Solutions	Times used	Success 0-3
Political and government challenges	Limited government capacity (includes poor systems, turnover, corruption)	10	Provide assessment services to employers when MOL inspections lacking	1	2
	Poor political will, political resistance to project, and low levels of government ownership of project activities	9	Narrow project activities to focus on most successful technical aspects	3	2
			Engage more with other partners	1	2
			Engage Ministry of Labor and seek MOU to promote sustainability	1	0
Be flexible with policy and reform proposals	1	0			
Political uncertainty, unrest, and turmoil; rapidly changing conditions	8	Be flexible with activities, focus on communication, and build partner HR	3	2	
Project deficiencies	Insufficient implementer capacity, including with monitoring	7	Bring in consultants	1	2
			Simplify indicators and improve measurement strategy	1	1
	Project team hiring difficulties and turnover	4	No solutions mentioned	N/a	N/a
	Limited project reach (in terms of geography or sector)	4	Redesign core project activities in a mobile format	1	2
			Lean on central project administration resources (e.g., from ILO)	1	2
	Poor project synergy with allied projects or parent organization	3	Hire national program coordinator ¹	1	.m
	Inadequate dosage (includes funding, intensity and duration of activities)	2	No solutions mentioned	N/a	N/a
Project contractual challenges (includes financial reporting requirements)	2	No solutions mentioned	N/a	N/a	
Poor representativeness of participants or insufficient worker outreach	2	Increase outreach and engagement activities	1	2	
Union, worker, and employer challenges	Low union or labor federation capacity (including workers lacking CB tools)	5	Co-conduct a strategic planning exercise	1	2
	Employer reluctance or low motivation to engage in the project	3	Engage through awareness raising, reframing efforts	1	2
	Workers experience threats or backlash as a result of project participation	3	Factories covered by Alliance and Accord systems were more responsive ²	1	2
			Invite union intervention to protect worker rights	1	3
			Work with labor federation and Accord backing to reinstate workers	1	2
	Low stakeholder understanding or limited public support for project efforts	3	No solutions mentioned	N/a	N/a
	Low levels of women's participation in project activities ³	2	Hold women-only activity sessions	1	1
	Family financial concerns place restrictions on worker participation	1	Support union leaders in working with participants' families to gain buy-in	1	1
Low levels of trust among stakeholders	1	Be flexible with stakeholders and emphasize strong communication	1	1	
Low beneficiary participation in project activities (e.g. in committees)	1	No solutions mentioned	N/a	N/a	
Other	Natural disaster (hurricane)	1	Follow security and emergency measures	1	0
	Price competition incentivizes buyers and employers to cut corners	1	No solutions mentioned	N/a	N/a
	Working with migrant labor (in delivering activities across language and culture barriers, with high worker turnover)	1	Encourage trade unions to communicate better with workers	1	.m
Total challenges		73	Total times solutions applied	26	

¹.m indicates the evaluation did not report the level of success of the solution. ² This solution was not applied by the project, instead coming from outside prior agreements. ³ More than two projects had low levels of women's participation, but only two evaluations clearly identified it as a challenge faced. CB = collective bargaining; HR = human resources; ILO = International Labor Organization; MOL = Ministry of Labor.

2. The most common challenges and solutions applied

In this section, we examine the five most common challenges that projects faced and explore illustrative examples of the most and least successful strategies projects deployed to mitigate them. We coded each solution as not successful, partially successful, moderately successful, or completely successful based on the degree to which the solution mitigated the challenge.

Challenge 1: Limited government capacity (includes poor systems, turnover, corruption)

Limited government capacity affected 10 of the 19 projects, and manifested as high staff or leadership turnover, corruption, and general inability to efficiently deliver services such as inspection or sanctioning of workers' rights violations. It is worth noting the challenge of limited government capacity largely went unaddressed; evaluators cited 1 instance in which implementers sought to mitigate the challenge. Table III.4 presents an example of a challenge and solution related to limited government capacity.

Table III.4. Example challenge and solution: Limited government capacity (includes poor systems, turnover, and corruption)

Challenge	Solution	Outcome
In the case of BW Jordan, the challenge of limited government capacity appeared in the Ministry of Labor's inability to provide inspection services. During and before the period in which the project took place (2014-2017), the ministry experienced high turnover of ministers, which in turn caused the labor inspectorate to be neglected. This neglect meant the inspectorate was not able to carry out basic inspections in RMG factories.	In the course of conducting compliance assessments, the BW Jordan implementer provided a degree of inspection services in place of the inspectorate. The assessments themselves did not directly rebuild the capacity of the labor inspectorate, but BW assessors also hosted three labor inspectors from the ministry who shadowed assessments and built their skills. Together, these solutions were moderately successful in addressing poor government inspection capacity.	Moderately successful 

Challenge 2: Poor political will, political resistance, and low government ownership of project activities

Strong political will was a key support for projects working with government agencies, but the challenges of poor political will, which included political resistance and low government ownership of project activities, appeared nine times.³⁴ Projects sought to address it six times, with mixed success. Three successful approaches and one unsuccessful approach are presented in Table III.5.

Table III.5. Example challenges and solutions: political will, political resistance, and low government ownership of project activities

Challenge	Solution	Outcome
The Mexico IGUALDAD project faced poor buy-in from certain state-level Ministry of Labor officials, whose unresponsiveness impeded strategic training workshops.	The implementer dropped that state from the intervention and focused greater attention on the remaining states. This was moderately successful in mitigating (or at least avoiding) issues with poor political will.	Moderately successful 

³⁴ Not all of these projects were primarily government-targeting. Some employer- and worker-targeting projects (such as Colombia WRC, described in Table III.5) also struggled with political resistance to their efforts and low government participation in relevant project activities.

Challenge	Solution	Outcome
The Bangladesh ILO F&BS project faced political resistance to interagency collaboration from government partners.	The implementer focused collaboration agreements only on technical and engineering areas, avoiding more sensitive political issues. This was moderately successful in mitigating political resistance.	Moderately successful 
The Colombia WRC project, which offered legal support to workers, faced a national political administration which, according to the evaluator, was “business friendly” and had created “a more challenging environment in which to advance workers’ rights.”	The implementer pursued strong and varied partnerships with supportive civil society organizations. This approach was moderately successful in substituting for the lack of government support.	Moderately successful 
In Haiti, the sustainability of the MCB project’s advances were threatened by the fact that the Ministry of Labor did not take ownership of continuing project activities.	The project implementer sought memoranda of understanding with relevant partners and directed a project task team to tackle the issue, but the steps were not successful in overcoming the challenge.	Not successful 
Also in Haiti, evaluators noted that MCB’s efforts relied on the improvement of labor inspections through the passage of a new labor law, which was at risk of delays in Parliament.	The project sought to “guarantee flexibility” to adapt to possible (but not assured) reforms in the labor code, but without an improved legal structure, such steps were not successful in increasing inspection effectiveness as planned.	Not successful 

Challenge 3: Political uncertainty, unrest, and turmoil; rapidly changing conditions

Political uncertainty, unrest, and turmoil, along with rapidly changing conditions, affected projects by causing delays.³⁵ These challenges occurred eight times, and projects applied solutions (all of which were moderately successful) three times. These solutions were maintaining flexibility with activities, focusing on communication with partners, and building partner human resources. Two of those cases are presented in Table III.6.

Table III.6. Example challenges and solutions: political uncertainty, unrest, and turmoil; rapidly changing conditions

Challenge	Solution	Outcome
The Georgia SWOG project faced rapidly changing project conditions, mostly in terms of political turmoil and a shifting legal landscape.	Georgia SWOG was moderately successful in addressing this challenge by directly providing human resources to project partners, stabilizing their staffing and helping them respond to changing environment.	Moderately successful 
The Egypt EEFMP project faced political uncertainty and low initial levels of trust between partners.	Egypt EEFMP was moderately successful in addressing this challenge by increasing communication with partners and maintaining flexibility with activities and timelines (flexibility demonstrated by both USDOL and the implementer).	Moderately successful 

³⁵ In some cases, such conditions sparked or fed into other challenges, such as poor political will or low government capacity.

Challenge 4: Insufficient implementer capacity of implementer

Evaluators cited insufficient implementer capacity across seven projects (most often with poor monitoring and evaluation systems), a challenge that implementers sought to mitigate only twice. Those cases are discussed in Table III.7.

Table III.7. Example challenges and solutions: Insufficient implementer capacity

Challenge	Solution	Outcome
The Bangladesh ILO F&BS project faced challenges with a subrecipient which did not have sufficient capacity to conduct and report on planned assessments, potentially setting up a bottleneck for project activities.	The project was moderately successful in overcoming this challenge by standardizing the assessment format and bringing on external engineering consultants to finish the assessments.	Moderately successful 
The Philippines BC-PLIP project used a poorly designed, overly complex monitoring and evaluation system which made tracking and planning activities difficult.	The project reduced the number of indicators in the monitoring system and sought to measure them better. This step was only partially successful , because the original deficiencies of the monitoring and evaluation system remained.	Partially successful 

Challenge 5: Low union or labor federation capacity

Evaluators cited low union or labor federation capacity as a challenge in five projects. This challenge overlaps with the motivation for several projects (in terms of developing union capacity) but is distinct in that it arose as an unforeseen problem, rather than a planned area for project activities. Table III.8 describes a challenge and solution related to low union or labor federation capacity.

Table III.8. Example challenges and solutions: Low union or labor federation capacity

Challenge	Solution	Outcome
A Georgia ICLLG partner was a trade union federation that had low capacity in both providing services to its affiliates and representing the labor movement to other sectors.	The project conducted a strengths, weaknesses, opportunities, and threats (SWOT) analysis with the federation leadership, which then allowed the project staff and labor leaders to engage in a strategic planning exercise which laid out actions to address the identified weaknesses. This solution was moderately successful in helping the federation adjust its structure and improve its capacity in service provision and representing the labor movement.	Moderately successful 

Recommendations for approaching common challenges
<ul style="list-style-type: none"> • Donors and implementers should design projects to be able to face one or more of the challenges listed in Table III.3, in particular: <ul style="list-style-type: none"> – limited government capacity – political will or resistance – political uncertainty or social unrest – insufficient implementer capacity – low union or labor federation capacity • Donors should prepare a toolkit of resources for implementers to draw from to mitigate the challenges, focusing on strategies that implementers found were successful in the past, such as: <ul style="list-style-type: none"> – Supporting low-capacity governments by providing temporary services that supplement poor services offered by official agencies

- Cutting unresponsive project partners or stakeholders and refocusing efforts on high-potential partners and stakeholders
- Narrowing partner agreements to a technical scope if political sensitivity is present
- Pursuing strong civil society, NGO, or private sector partners if government partners are unsupportive
- Providing human resources (such as legal expertise) to partners enduring political turmoil or rapidly changing project conditions
- Increasing communication with partners and maintaining flexibility with activities to overcome political uncertainty and mistrust among stakeholders
- Addressing inadequate implementer capacity by bringing in external consultants and/or working with partners to transition ownership of project-initiated activities to local stakeholders
- Preparing to guide low-capacity union partners through extensive strategic planning processes
- Preparing for risks related to potential backlash (from employers or community members) against project participants
- Using the following tools during project conception and design:
 - o Needs and opportunities assessments
 - o Risk analysis systems
 - o Stakeholder analysis tools

This page has been left blank for double sided copying.

IV. Conclusions and recommendations

This report synthesized the findings of evaluations of 19 projects that aimed to build government inspection and enforcement capacity, support employer compliance with labor laws, and or build worker and union capacity for advocacy. Although the nature of this review precludes us from identifying causal relationships between project characteristics and outcomes, our review revealed relationships between some project features and project effectiveness, recurring challenges that OTLA and implementers should prepare for, and opportunities for improving project data practices moving forward.

In this chapter, we begin by summarizing recommendations evaluators made in individual project evaluations in Section A. In Section B, we review the synthesis review’s key findings, discuss their implications, and provide recommendations for OTLA on setting priorities for future projects, strengthening support for project implementers based on the synthesis’s findings, and how to structure the requirements for evaluators moving forward.

A. Recommendations from evaluations

Each of the 19 evaluations presented recommendations for improving the performance, effectiveness, and sustainability of the project at hand (if ongoing) or future projects. We consolidated the numerous suggestions into 11 main recommendation groups in the following table.

Table VI.1 Evaluator-provided recommendations

Number	Recommendation group	Number of projects where recommended
1	Combine, expand, or realign project activities to better fit needs ³⁶	17
2	Improve M&E systems, project PMP, and/or accountability requirements	14
3	Provide greater assistance and resources to low-capacity government partners	13
4	Revise sustainability plan or embed sustainability into project design	11
5	Help partners engage one another more, open greater communication	8
6	Clarify project priorities, structure, and roles for staff and partners	6
7	Adjust budget, allocate funds more clearly, or set up outcome-based budgeting	4
8	Increase funding / streamline disbursement of funds (particularly in ILO projects)	3
9	Provide more strategic, targeted assistance to unions or employers	3
10	Improve gender-conscious programming	3
11	Showcase or disseminate project successes	2

Most of the recommendations offered by evaluators align with our recommendations, which are presented in the following subsection.

³⁶ This group includes recommendations such as “Combine Fire and Building Safety and OSH” (*combine*) or “Expand EEFMP to all governorates” (*expand*) or “Ensure that project priorities are clearly reflected in the project’s logic model so that the project monitoring plan (PMP) and the work plans better reflect the project’s goals” (*realign*). Each recommendation in this group relates to delivering interventions that are comprehensive, right-sized, and linked to core goals and beneficiary needs.

B. Conclusions and recommendations derived from synthesis of evaluations

Our recommendations are derived from the analysis of 1) the project design factors associated with project effectiveness (including USDOL-influenced, implementer-controlled, and contextual factors), 2) project challenges and solutions, and 3) evaluation reports and data quality. The conclusions and recommendations offered in this section are synthesized from the findings presented throughout the report. These conclusions and recommendations represent our assessment of patterns and trends in projects' level of effectiveness in achieving inputs, outputs, and outcomes and factors that may influence their effectiveness as well as our qualitative analysis of evaluators' comments on project implementation. The trends we have identified may be based on real relationships; however, given that 1) our analysis is based on a finite sample of evaluations of 19 projects and 2) lower-quality evaluations tended to report higher project effectiveness, there is a possibility that the relationships we have described in this report and highlighted in this section may instead be due to chance or overstated findings. Nonetheless, we consider these findings worth taking into consideration as USDOL moves forward making decisions about how to prioritize scarce resources and support those implementing important projects intended to support workers' rights around the world while promoting an even playing field for workers in the United States. We list project strategies used by the most effective projects in Chapter III, Section C, sub-section 2.

1. Conclusions and recommendations for project priorities and design factors related to project effectiveness

Projects with more comprehensive scopes and goals appear to be more effective.³⁷ Several of our findings have a common theme: projects that are more comprehensive, or are part of more comprehensive efforts, are more likely to succeed. **We found that projects with larger budgets, longer periods of performance, and fully tripartite approaches were more likely to meet their goals than shorter projects with smaller scope.** Similarly, projects that were aligned with ongoing activities in the target country had greater success. These findings suggest that OTLA may wish to prioritize such large-scale, comprehensive projects moving forward.

Projects primarily targeting government stakeholders generally had the lowest effectiveness scores of the 19 projects while those primarily targeting workers, unions, and labor federations tended to have greater effectiveness in meeting project goals. Projects' objectives of improving labor law compliance were more closely aligned with the interests of unions and employers while governments were more likely to have been constrained by competing priorities, scarce resources, and political instability.

Implementer capacity plays a key role in project effectiveness, suggesting a potential role for expanded guidance for implementers. Projects with the most coherent, evidence-based theories of change were among the most effective (see Appendix C for guidance in this area). Similarly, projects with high-quality management structures and high efficiency were more likely to meet their goals, using effective strategies to overcome challenges such as low political will or weak government capacity.

³⁷ Projects that had overly ambitious goals in their PMPs (ones that were not calibrated to what was feasible given time and budget) indeed fell short of them. However, projects that were comprehensive in their scope, as in ambitious in their partnership building or in their comprehensive delivery of services to a given sector, tended to do better in meeting their goals. Projects that sought to approach *multiple actors and aspects* of a given sector were more successful than those who were more focused on a single actor or aspect of the sector.

Weak monitoring and evaluation systems and PMPs hindered projects in achieving, tracking, and reporting their progress. For those 10 projects with poorly designed PMPs, implementers were less able to convey whether they had met their quantifiable goals, and how the provision of inputs may have produced outputs or achieved outcomes. This limitation also hindered the ability of evaluators to draw lessons from the projects in terms of effective activities and the efficiency of the project overall. Projects with stronger monitoring and evaluation systems and PMPs could identify the specific areas of success in their interventions, adjust strategies, and improve implementation mid-course, while evaluators of those projects were able to assess the effectiveness of project components and their contributions to the ultimate goals of the interventions.

Projects improved gender equality outcomes when gender was part of project design. As we discussed in Box III.3, only two projects set gender equality as a core objective, but six projects reportedly achieved positive outcomes in terms of gender equality or women’s empowerment in the workplace. Evaluators considered mainstreaming gender as a cross-cutting issue in project design to promote women’s equality outcomes.

Projects’ strategies have supported vulnerable populations with moderate success, but we can conclude little with only three projects supporting vulnerable populations. As described in Box III.4, three projects targeted vulnerable groups such as migrants or illegally contracted workers. These projects used diverse strategies (such as supports for legal action or unified contracts) and each had success in advancing protections for vulnerable groups.

The findings discussed above motivate the following set of recommendations, which we present for USDOL and for implementers. USDOL may consider providing guidance to implementers to support them in carrying out the recommendations for implementers.

Recommendations for USDOL

- Consider funding **longer-term interventions with appropriately sized budgets** to achieve maximum progress toward project goals.
- Consider increasing the share of OTLA projects where the **primary target group is workers**, unions, or their federations, given these projects appear to be relatively low in cost and relatively high in effectiveness.
- Consider investing in projects (and advocating for project activities) that **align with similar efforts in the target country** to produce synergies and expand impact.
- To **develop evidence-based theories of change**, dedicate necessary resources before the solicitation stage to locate evidence related to the project objectives and examine the proposed interventions against that evidence. To facilitate repeated reviews of the evidence, USDOL may wish to compile evidence of the effectiveness of various specific intervention activities that are commonly part of USDOL-funded projects in an organized or searchable document that could be accessed when conducting evidence reviews for the theory of change of any new project. This document should be updated regularly to incorporate new evidence.
- Consider **selecting implementers that have demonstrated capacity** to execute all project components or be prepared to provide initial and ongoing guidance and support for implementers in essential areas, such as monitoring and evaluation.

- Consider developing and sharing with implementers a **toolbox of strategies to mitigate challenges** with political will and government capacity in government-targeting projects (see *Recommendations related to common challenges and strongest solutions*, below).
- Consider requiring projects to **mainstream gender considerations** into their designs to improve gender equality outcomes.
- Require **strong monitoring and evaluation, PMP, and accountability systems** and support implementers in these efforts with written resources and mentoring.
- Consider requiring implementers to develop a **risk management** plan as one of their first steps.
- Prioritize **flexibility** with implementers as they combine, expand, narrow, or realign core project activities to address evolving risks and persistent challenges.
- Consider committing **more funding to projects facing high risks** and/or (in the case of ILO projects) consider using a streamlined, “quick-release” disbursement mechanism to deliver funds to the implementer in a timely manner.

Recommendations for implementers

- At project award, **refine an evidence-based theory of change** with donors and stakeholders to support effectiveness (see Recommendations for USDOL above and Appendix C for more guidance).
- **Design for sustainability, then (after award) draft, follow, and regularly update a sustainability plan** to increase the likelihood that project impacts are maintained.
- **Design for flexibility** in combining, expanding, narrowing, or realigning core project activities to address evolving risks and persistent challenges, while staying within the original scope of the grant.
- **Anticipate risks** by developing a risk management plan as one of the first steps of the project.
- Reach out and **offer additional training and technical assistance** to low-capacity target governments.

Recommendations related to contextual factors

The presence of contextual factors such as economic or political conditions does not appear to have a strong, direct relationship with project success. This is likely because some projects adopted strategies to address them, such as increasing communication with reluctant government actors. This implies projects should pay close attention to contextual factors that could delay or impede activities—especially government capacity and political will—and deploy solutions early that could mitigate the effects of challenging conditions. Such solutions are listed in the following section (Recommendations related to common challenges and strongest solutions) as part of a toolbox USDOL could prepare for implementers.

2. Recommendations related to common challenges and strongest solutions

Challenges confronted every project, and implementers applied solutions with mixed success. By drawing lessons from successful and unsuccessful solutions, we can offer recommendations for strategies that could help future projects overcome similar challenges.

First, donors and implementers should design projects to be able to overcome one or more of the challenges listed in Table III.3, particularly those that arose most frequently:

- Limited government capacity
- Political will or resistance
- Political uncertainty or social unrest
- Insufficient implementer capacity
- Low union or labor federation capacity

Second, to support the resilience of projects once they are running, USDOL could prepare a toolbox of resources for implementers to draw from as they seek to mitigate common challenges. The toolbox could include a set of how-to briefs for implementers focusing on solutions that were successful in the past, such as:

- Supporting low-capacity governments by providing temporary services to supplement poor state-offered services while also building their capacity (for example by allowing government officials to partner in implementation) with a plan to transfer ownership of the tasks before project close.
- Cutting unresponsive project partners and stakeholders and refocusing efforts on high-potential partners and stakeholders.
- Narrowing partner agreements to a technical scope if political sensitivity is present.
- If government stakeholders are unsupportive, pursuing partnerships with civil society, NGOs, the private sector, or focusing efforts on stakeholders showing least resistance as appropriate.
- Providing human resources (such as legal expertise) to partners enduring political turmoil or rapidly changing project conditions.
- Increasing communication with partners to address mistrust among stakeholders, facilitated by developing a Stakeholder Management Plan early on.
- Maintaining flexibility with activities to overcome political uncertainty and mistrust among stakeholders.
- Addressing inadequate implementer capacity by bringing in external consultants and/or working with partners to transition ownership of project-initiated activities to local stakeholders.
- Preparing to guide low-capacity union partners through extensive strategic planning processes.
- Preparing for risks related to potential backlash (from employers or community members) against project participants.
- Using the following tools during project conception and design (both USDOL and implementers):
 - Needs and opportunities assessments
 - Risk analysis systems
 - Stakeholder analysis tools

3. Recommendations related to evaluations and data quality

Evaluations reviewed for this report were mixed-methods, non-causal performance evaluations, and were generally medium-quality. Most of the qualitative data sources were subject to bias while the quantitative data provided through implementers' PMPs were not consistently reported. The validity of evaluation findings was also threatened by the timing of the evaluations because some evaluations were published before projects ended. The following recommendations are designed to help USDOL,

evaluators, and implementers avoid those evaluation and data quality deficiencies. We recommend that USDOL emphasize clear expectations on data reporting requirements expected of implementers and provide guidance to optimize the quality of data collected and used for projects' evaluations. The following are specific recommendations for evaluators, USDOL, and implementers.

Recommendations for evaluators and USDOL

- **Engage evaluators at the project design stage** (preferably pre-award) to support the development of the most rigorous evaluation design possible.
- **Require evaluations to collect data at project end—not before** to allow for clearest view of project success.
- Consider supporting evaluators in **conducting follow-up data collection** 1 year after project completion to measure longer-term outcomes.
- Provide evaluators with the **resources and time to conduct fieldwork that captures a full view** of the project.
- Encourage evaluators to **supplement qualitative data with quantitative data from rapid scorecard or small surveys**.
- Support evaluators in **developing a representative sampling frame** of potential interviewees and ask implementer to help with outreach (not final selection) to those respondents.
- Ensure that evaluators use **interactive, participatory validation workshops** (virtually, if need be) with stakeholders to review findings and collect any clarifying information to improve evaluation accuracy, as required by USDOL.
- Include in evaluator terms of reference (TORs) the requirement that evaluators present and dissect **projects' results frameworks, theories of change, assumptions, challenges, and solutions** in their evaluations.
- Include in evaluator terms of reference (TORs) the requirement that evaluators assess **projects' gender approaches and impacts, as well as any changes in stakeholder policies, programs, or resource allocation** as a result of project activities.
- **Upgrade the required PMP reporting template to an easy-to-use electronic platform** and require its use by all implementers to make data transparent, comparable, and useful for reviewers synthesizing trends across projects and reduce paper-based data collection and reporting.
- **Consider developing a set of standard indicators to measure long-term outcomes across all projects or all projects with similar long-term objectives**. Such a set of indicators would facilitate cross-project comparisons.
- **Make project budgets and financial reports available to evaluators** (and add the review of such items to evaluation Terms of Reference) to ensure evaluators can assess project efficiency.

Recommendations for implementers

- **Use an upgraded electronic PMP reporting template**, provide background and justifications for all indicator targets, and report all indicators on a semi-annual basis to ensure data availability and project accountability.

- **Disaggregate all relevant indicators by gender** to improve tracking of this important cross-cutting issue.
- Ensure that **each output, short-term outcome, and other item in the results framework has indicators established and targets set** (see Appendix C for guidance in this area).
- **Adhere to the following indicator target-setting best practices:**
 - 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.**
- Use **outcomes-based budgeting systems**, keep them updated, and have regular budgeting check-ins between USDOL and implementers for assessments of efficiency and calculations of cost per outcome in the project results framework.

By documenting and synthesizing past challenges and solutions, USDOL can offer implementers ready-made strategies to mitigate common threats and deliver the most successful projects possible.

This page has been left blank for double sided copying.

Appendix A:

List of documents used

This page has been left blank for double sided copying.

Table A.1. Projects and evaluations included in the synthesis review

Project	Evaluator	Evaluation name	Date published	Evaluation manager
Bangladesh (BWB)	OAI	Independent Evaluation of the U.S. Department of Labor's Technical Cooperation Portfolio to Promote Workers' Rights in Bangladesh	10/2015	USDOL
Bangladesh (SC F&BS)	OAI	External Final Evaluation of the Improving Fire and Building Safety for Bangladesh's RMG Workers Project	9/2018	USDOL
Bangladesh (ILO F&BS)	Jonathan Price	Final Independent Evaluation of Improving Fire and General Building Safety in Bangladesh	12/2016	ILO
Colombia (CAL)	IMPAQ	Interim Performance Evaluation Workers' Rights Centers for the Greater Protection of Labor Rights in Colombia	10/2019	USDOL
Colombia (PCILSC)	OAI	Promoting Compliance with International Labor Standards in Colombia: Independent Final Evaluation	1/2017	USDOL
Egypt (EEFMP)	Universalia Management Group	Independent final evaluation of the Egypt Export Factory Monitoring Project	9/2018	ILO
Georgia (SWOG)	MSI	Final Performance Evaluation: Improving Compliance with Labor Laws in Georgia and Strengthening Workers' Organizations in Georgia	2/2019	USDOL
Georgia (ICLLG)	Katerina Stolyarenko	Final Independent Evaluation of ICLLG project	7/2019	ILO
Haiti (MCB)	Juan-David Gonzales	Final evaluation of "Building Capacities of the MAST to Ensure Labour Law Compliance in the Haitian Apparel Sector"	12/2017	ILO
Haiti (BWH)	IMPAQ	Interim Performance Evaluation: Better Work Haiti	10/2019	USDOL
Jordan (BWJ)	Amy Jersild	Better Work Jordan: Phase 2 Final evaluation	10/2018	ILO
Malaysia (LL-IRR)	OAI	An Independent Multi-Project Evaluation of Protecting the Rights of Migrant Workers Through Empowerment and Advocacy in Malaysia and Support for Labor Law and Industrial Relations Reform in Malaysia	10/2018	USDOL
Mexico (IGUALDAD)	OAI	Independent Final Evaluation of the Initiative to Guard Against Labor Discrimination Project "Igualdad"	4/2018	USDOL
Morocco (Wad3éyati)	Sistemas, Familia, y Sociedad	Independent Final Evaluation: Wad3éyati, Gender Equality in the Workplace in the Kingdom of MOROCCO	10/2017	USDOL
Peru (PLIP)	IMPAQ	Final Performance Evaluation: Building the Capacity of the Peruvian Labor Inspectorate & Building Union Capacity to Reduce Precarious Employment in Peru	8/2019	USDOL
Peru (BUCCPEP)				
Philippines (BC-PLIP)	IMPAQ	Final Performance Evaluation Building the Capacity of the Philippines Labor Inspectorate	9/2019	USDOL
Vietnam (ILO-IR II)	Katerina Stolyarenko	Final Independent Evaluation of ILO-IR Phase II project (2012-2016)	9/2016	ILO
Vietnam (BWV-UCD)	OAI	Independent Multi-Project Evaluation: Improving Labor Laws and Labor Administration Within the New Industrial Relations Framework (NIRF) and Better Work Vietnam, Union Capacity Development Component (BWV-UCD)	7/2018	USDOL

This page has been left blank for double sided copying.

Appendix B:

Detailed description of analysis approach

This page has been left blank for double sided copying.

I. Detailed description of analysis approach

The analysis approach for this synthesis review involved extracting key information from each evaluation into a rubric, coding those rubric items to categorical variables, extracting performance information from project PMPs, assessing project effectiveness using a systematic scoring system, and exploring the relationships between project categorical variable values and project effectiveness. We then reviewed the rubric and the evaluations to flesh out the relationships identified between project factors and effectiveness. Finally, we grouped project challenges and solutions identified in each evaluation and assessed the success of each solution in mitigating the challenge. We provide more detail on each step of this process here.

A. Extracting information to the rubric

The data extraction process involved detailed information from each of the 19 evaluations into a 61-item rubric. These items included:

- **Project characteristics** (per evaluators' assessments), such as funding, implementer, duration, economic indicators, trade relationships, project focus, target groups, inputs, outputs, immediate outcomes, ultimate outcomes, and theory of change;
- **Evaluation features** (per reviewers' assessments), such as data validity, analytical techniques, limitations, data sources, overall evaluation quality, and publication date relative to project end;
- **Project implementation** (per evaluators' assessments), such as management and timeline, design strengths and weaknesses, implementer capacity, budget, efficiency, communication structures, project targets, PMP quality, effectiveness in meeting project goals, *how* those goals were (not) achieved, risk management, and sustainability plans;
- **Project outcomes** (per evaluators' and reviewers' assessments), such as changes in policy, gender equality, vulnerable groups' rights, the legal framework, enforcement, compliance, partner resource allocation, systems change, and evaluator recommendations.
- **Contextual factors** (per evaluators' and reviewers' assessments), such as political climate, economic conditions, prevailing attitudes toward workers' rights, and the interest and influence of stakeholders
- **Project challenges and solutions** (per evaluators' and reviewers' assessments), including each challenge that each project faced and how solutions (if any) were applied.

Each of these rubric items was populated with notes from our interpretation of the evaluations and quotes extracted directly from the evaluations, with enough detail to inform categorization (the next step in our analytical approach).

B. Coding rubric contents to categorical variables

Using the contents of the rubric, we broke each rubric item into variables which could be scored with binary or categorical values. For example, for the rubric item *Factors or opportunities that will contribute to project success (during life of project)*, we produced four columns, each headed by a supporting factor group, such as *alignment with other projects or efforts in the country*, and gave each project a "1" if it had that supporting factor and a "0" if it did not have it. For variables where we needed to capture the *degree* to which something was true for a project, we used an ordinal scale (from 0 to 2 or 0 to 3). For example,

for the rubric item *project management and timeline*, we extracted two variables: quality of project management, scored from 0 (very poor) to 3 (very good), and project delays, scored from 0 (none) to 3 (severe).³⁸ Evaluation characteristics were scored using the same system—for example, the severity of the evaluation’s limitations or the degree of data validity the evaluation had.

When evaluations did not report details for a certain rubric item—for example, project efficiency—we gave a .m (missing) value for that project in that categorical variable. For variables that were not relevant for certain projects, such as whether a solution was applied (when the relevant challenge was never present), we gave n/a. In general, .m values indicated a lack of attention on the part of the evaluator, while n/a simply indicated the item was not relevant to the project. Most evaluations had only a handful of missing items, but several had more than 25 missing variable values. When the necessary detail was missing from the rubric items to code key variables, we were forced to exclude those projects from those areas of analysis. For example, one evaluation completely lacked information on project management and efficiency, and we had to exclude that project from our examination of how management and efficiency relates to other project characteristics and project effectiveness.

Next, we added composite and summary variables to capture important trends not directly categorized from the rubric, such as *number of supporting factors* each project had and *number of gender-disaggregated indicators tracked in PMP*. These variables aided our interpretation of associations between project factors and effectiveness and helped identify gaps in the evaluations and the project PMPs.

Coding rubric items to categorical variables, adding composite and summary variables where necessary, and scoring projects for effectiveness (see next section) produced 238 variables. Manually coding 19 projects to those variables produced 4,522 values, all of which were examined in the analysis stage.

C. Effectiveness scoring

Given problems with PMP data quality identified across most projects and described in Section III.A, we could not use the results-to-goals (RGR) approach that we proposed in the Methodology Plan to assess project effectiveness. Instead, we assessed effectiveness using a comprehensive view of projects as described in the evaluations and an A-B-C scoring system.³⁹

Under the A-B-C scoring system, projects receive scores of A, B, or C for the degree to which they met goals for inputs, outputs, and outcomes.⁴⁰ Projects could receive different scores in different areas (e.g. an A for inputs, an A for outputs, and a B for outcomes). Scores were assessed as follows:

³⁸ Variables differed in direction, i.e. sometimes 0 was a desirable score (as with severity of delays), and sometimes 0 was the worst score (as with project management quality). However, we tracked the direction of the variable in the analysis stage to ensure correct interpretation of results.

³⁹ Some evaluations did not explicitly state the planned inputs or outputs for their respective projects, and some projects did not report their concrete indicator goals for those elements in their PMPs. In those cases, we closely read the evaluations to identify implied activities and outputs.

⁴⁰ We originally scored inputs, outputs, and outcome achievements using a 0-3 scale for ease of data manipulation. No projects earned 0 values. 1 scores received a C, 2 received a B, and 3 an A. This allowed us to report scores in both an intuitive grade system and manipulate the number scores to look for patterns in the data. We also summed scores for inputs, outputs, and outcomes for a total score which allowed for high-level comparison. Summed scores less than 5 received a C, sums of 5, 6, and 7 received a B, and those scoring 8 or higher received an A. With this

- Projects which did not deliver their core planned inputs, outputs, or outcomes received a C input, output, or outcome score respectively.
- Projects which delivered some of their core planned inputs, outputs, or outcomes received a B input, output, or outcome score respectively.
- Projects which delivered all (or nearly all) of their core planned inputs, output, or outcomes received an A input, output, or outcome score respectively.

Input, output, and outcome scores were accompanied by two other letter scores that were discussed in Chapter III, Section C of this report: *quality of project management and efficiency* and *the level of partner and beneficiary buy-in*. The project management and efficiency scores were computed by taking the mean of project management, operational budget efficiency, and project efficiency values as assessed by the reviewer. This mean was then coded into an A, B, or C score depending on its value. The partner and beneficiary buy-in scores were computed by taking the mean of a supporting factor score (specifically, the level of buy-in/enthusiasm of beneficiary/participant staff), a design strength score (the degree to which the project capitalizes on political will), and the level of private sector and government buy-in, as assessed by the reviewer. This mean was then coded into an A, B, or C score depending on its value.

D. Identifying relationships with effectiveness

To analyze the relationships between project factors and A-B-C scored items, we examined trends and patterns that became visibly apparent in tables sorted to isolate projects with specific characteristics.⁴¹ In Figure B.1, we present a collapsed, example version of the analytical table used to identify associations between project characteristics and effectiveness. In this example, the principal target group column is sorted, allowing the researcher to examine patterns among specific target groups in terms of project management and efficiency, level of partner buy-in, and the effectiveness of projects in achieving inputs, outputs, and outcomes. Using this sorting system, it is visually apparent that projects where government ministries were the primary targets had lower effectiveness scores than fully tripartite or worker-focused projects.

system, each overall letter score was also the mode of the input, output, and outcome scores. For example, a project with an A for inputs, A for outputs, and B for outcomes received an overall score of A. The last two columns in the screenshot of presented in this section show the overall scores (letter and number) for each project.

⁴¹ There is a risk that by using only three buckets (A, B, and C), the project scores may have gravitated toward the center of the scale. However, the A-B-C system is the level of precision we are comfortable with, because further gradations imply we can detect minute differences between differently-evaluated and differently reported project aspects. Given the wide variation in the quality and completeness of the evaluations examining the projects, we cannot precisely say which project should be scored a 6 of 10, for example, or a 7 of 10. Larger scoring buckets allow for meaningful grouping without near-arbitrary assignment of narrow scores.

Figure B.1. Screenshot of analytical table

Project	#	Principal target group	Region	Total project value (million USD)	Performance period (year)	Quality of project management and level of efficiency	Level of partner and beneficiary buy-in	Degree to which inputs were successfully delivered	Degree to which outputs were successfully produced	Degree to which outcomes were successfully achieved	Overall score (see function)	Sum of input, output, and outcome scores
1. Bangladesh BWB	1	Employer	South Asia	4.80	2.3	C	B	B	B	C	B	5
11. Haiti BWH	11	Employer	LAC	13.30	10.8	A	C	B	B	B	B	6
12. Jordan BWJ Phase II	12	Employer	MENA	4.10	3.0	A	B	B	B	C	B	5
16. Morocco Wad3éyati	16	Employer	MENA	1.25	4.2	C	A	C	C	C	C	3
20. Vietnam BWV-UCD	20	Employer	SE Asia	2.30	6.0	A	B	A	A	B	A	8
3. Bangladesh ILO F&BS	3	Ministry	South Asia	1.50	3.2	B	C	B	B	B	B	6
5. Colombia PCILSC	5	Ministry	LAC	10.32	5.5	B	B	B	B	B	B	6
8. Georgia ICLLG	8	Ministry	Eurasia	3.00	6.3	B	B	B	B	B	B	6
10. Haiti MCB	10	Ministry	LAC	2.40	3.8	C	B	C	B	C	C	4
14. Malaysia LL-IRR	14	Ministry	SE Asia	1.60	3.5	C	B	C	C	X	C	2
15. Mexico IGUALDAD	15	Ministry	LAC	1.40	3.3	C	C	B	C	B	B	5
17.1. Peru PPIP	17.1	Ministry	LAC	2.00	4.5	C	B	B	B	B	B	6
6. Egypt EEFMP	6	Tripartite	MENA	10.00	6.4	A	B	B	A	A	A	8
18. Philippines BC-PLIP	18	Tripartite	SE Asia	1.25	4.7	C	A	B	B	A	B	7
19. Vietnam ILO-IR II	19	Tripartite	SE Asia	3.67	3.9	B	A	C	B	A	B	6
2. Bangladesh SC F&BS	2	Worker	South Asia	1.20	5.5	A	B	A	B	B	B	7
4. Colombia WRC-CAL	4	Worker	LAC	1.75	4.0	A	B	A	A	B	A	8
7. Georgia SWOG	7	Worker	Eurasia	0.94	4.0	X	B	A	A	B	A	8
17.2. Peru BUCCPEP	17.2	Worker	LAC	1.00	2.0	C	B	B	B	B	B	6

Where feasible, we also examined multiple project factors that were relevant to one another together before examining them against the input, output, and outcome scores. For example, we noted that project duration was associated with budget size, which in turn was associated with management quality and efficiency. This allowed us to identify not only the project strengths associated with effectiveness, but also the preceding factors associated with those strengths.

After identifying associations, we returned to the original detail-rich rubric to extract examples of the relationships between variables and flesh out the relationships in our narrative of the findings.

E. Examining challenges and solutions

Projects faced 73 individual challenges, which we grouped into 21 broader challenges, such as *project team hiring difficulties and turnover*. We then examined the strategies applied by projects to each challenge and scored the success of the solution in mitigating the challenge from 0 (not successful) to 3 (completely successful) and presented the resulting table in section III.D. We supplemented the table with an analysis of the top five most common challenges and the most and least successful solutions applied to each.

Appendix C:

Theory of change and indicator guidance

This page has been left blank for double sided copying.

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).⁴² 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 approaches⁴³ 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:

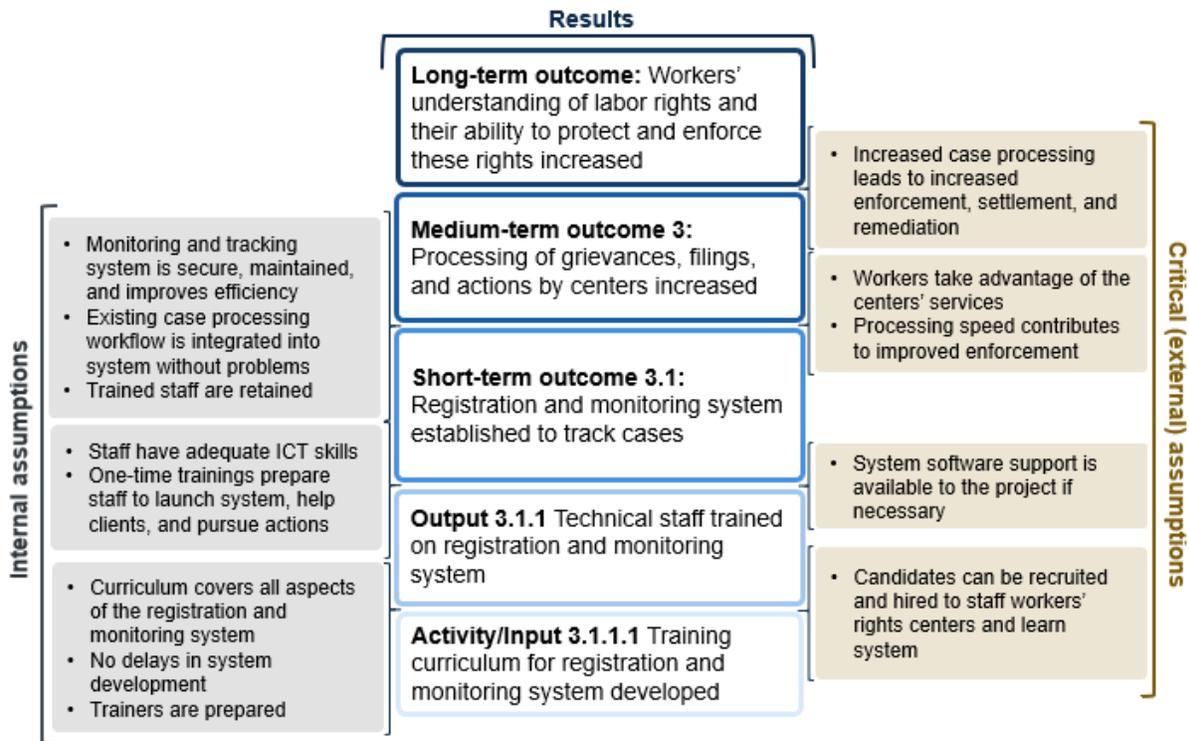
⁴² ILAB has indicated that it draws broader scopes at project conceptions, so that there is room to later shift and incorporate work that is deemed necessary to achieve the project objective. If scope is too narrow, leaves little to no room to elaborate results as implementation proceeds based on evolving 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.

⁴³ 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	<p><i>Number and percentage of inspectors trained on the new data system</i></p> <p>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.</p>	<ul style="list-style-type: none"> • <u><i>Number of inspectors trained on the new system</i></u> • <u><i>Percentage of all inspectors trained on the new system</i></u> <p>These indicators are direct because they are each particular to one data point and they clearly measure the intended result.</p>
2	<p><i>No. of inspectors <u>using</u> new data system</i></p> <p>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?</p>	<p><i>Number of inspectors <u>who have logged onto the data system at least two times in each of the previous six months.</u></i></p> <p>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.</p>

Indicator	Poorly defined indicator	Improved indicator
3	<p><i>Number of <u>new/revised procedures/tools</u> with support of the project <u>used</u> by labor and fire inspectors <u>and</u> privately contracted monitors supporting the National Initiative</i></p>	<p><i>Number of:</i></p> <ul style="list-style-type: none"> <i>new procedures</i> <i>revised procedures</i> <i>new tools</i> <i>revised tools</i> <p><i>which were developed with support of the project and are <u>used weekly by the majority of:</u></i></p> <ul style="list-style-type: none"> <i>labor inspectors</i> <i>fire inspectors</i> <i>privately contracted monitors supporting the</i> <i>National Initiative</i>
	<p>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.</p>	<p>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.”</p>
4	<p><i>Number of government <u>entities</u> that are charged with the investigation <u>and/or</u> prosecution of persons or groups that commit crimes with an anti-union motive that <u>adopt project materials</u> as part of their <u>internal trainings</u>.</i></p>	<p><i>Number of <u>national government entities</u> that are charged with the:</i></p> <ul style="list-style-type: none"> <i><u>investigation</u></i> <i><u>prosecution</u></i> <p><i>of persons or groups that commit crimes with an anti-union motive that <u>include</u> project:</i></p> <ul style="list-style-type: none"> <i><u>Case management system</u></i> <i><u>Labor law toolbox</u></i> <p><i>as part of their internal:</i></p> <ul style="list-style-type: none"> <i><u>onboarding</u></i> <i><u>ongoing trainings</u>.</i>
	<p>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.</p>	<p>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.</p>

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.⁴⁴ 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 (<i>sustained capacity and linkages</i>)
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 (<i>sustained resources</i>)
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 (<i>sustained motivation</i>)
Long-term outcomes
Worker (sustainability-oriented indicator): Percent of workers in targeted factories that indicate interest in future trainings from union leadership (<i>sustained demand</i>)
Government (<i>sustainability-oriented indicator</i>): Number of inspections pre-approved for next FY (<i>sustained service delivery</i>)
Employer (sustainability-oriented indicator): Percent of assessment costs covered by international buyers' subscriptions (<i>sustained access and demand</i>)

⁴⁴ 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>

This page has been left blank for double sided copying.

This page has been left blank for double sided copying.



Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom



mathematica.org