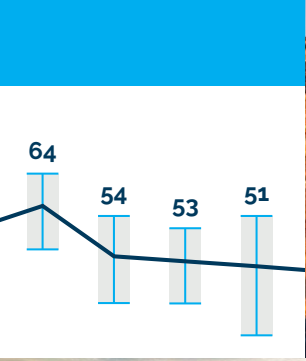


Results and Performance of the World Bank Group

AN INDEPENDENT EVALUATION



2020



WORLD BANK GROUP
World Bank • IFC • MIGA

© 2020 International Bank for Reconstruction and Development / The World Bank
1818 H Street NW
Washington, DC 20433
Telephone: 202-473-1000
Internet: www.worldbank.org

ATTRIBUTION

Please cite the report as: World Bank. 2020. *Results and Performance of the World Bank Group 2020*. Independent Evaluation Group. Washington, DC: World Bank.

COVER PHOTO

shutterstock

EDITING AND PRODUCTION

Amanda O'Brien

This work is a product of the staff of The World Bank with external contributions. The findings, interpretations, and conclusions expressed in this work do not necessarily reflect the views of The World Bank, its Board of Executive Directors, or the governments they represent. The World Bank does not guarantee the accuracy of the data included in this work. The boundaries, colors, denominations, and other information shown on any map in this work do not imply any judgment on the part of The World Bank concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

RIGHTS AND PERMISSIONS

The material in this work is subject to copyright. Because The World Bank encourages dissemination of its knowledge, this work may be reproduced, in whole or in part, for noncommercial purposes as long as full attribution to this work is given.

Any queries on rights and licenses, including subsidiary rights, should be addressed to World Bank Publications, The World Bank Group, 1818 H Street NW, Washington, DC 20433, USA; fax: 202-522-2625; e-mail: pubrights@worldbank.org.

Results and Performance of the World Bank Group

AN INDEPENDENT EVALUATION



2020

NOVEMBER 30, 2020



WORLD BANK GROUP
World Bank • IFC • MIGA



Contents

Abbreviations	iv
Acknowledgments	v
Overview	vi
1. Introduction	2
2. Part I: Assessing Performance through Ratings	6
World Bank Projects	7
Country Programs	12
Explaining the Trends	13
Responding to COVID-19 and Other Shocks	21
IFC Projects	24
Explaining the IFC Trends	27
MIGA Projects	33
3. Part II: Assessing Outcome Levels	36
Introduction	36
Outcome Classification Framework	37
Project Outcomes	42
Project Outcome Levels and Ratings	46
Thematic Area Outcomes	53
4. Conclusions: Getting to Outcomes	56
Findings and Conclusions	56
Implications	58
Looking Ahead	61
Bibliography	62
Photo Credits	64

Boxes

Box 1.1. Key Terms in This Report	3
Box 2.1. Aspects of Bank Performance	8
Box 2.2. Elements of Monitoring and Evaluation Quality	11
Box 2.3. Smoothing Country Program Ratings	13
Box 2.4. IFC's Reforms to Strengthen Upstream Engagement	31
Box 3.1. The Outcome Level Framework	40
Box 3.2. IFC's AIMM System for Setting Project Objectives	45
Box 4.1. Setting Project Objectives	57
Box 4.2. The Coronavirus Pandemic and IFC Project Ratings	59
Box 4.3. A Fresh Approach to Understanding Country Outcomes	61

Figures

Figure 1.1. Outcome Levels Classification	4
Figure 2.1. World Bank Project Outcome Ratings, Annual	7
Figure 2.2. Project Outcome Ratings, FY12–14 and FY17–19	8
Figure 2.3. Country Program Outcome Ratings, FCV and Non-FCV Countries	12
Figure 2.4. Decomposing the World Bank Project Rating Increase over FY12–14 and FY16–18	14
Figure 2.5. Outcome Rating Plotted against M&E Quality Rating	16
Figure 2.6. Ratings for Country Program Objectives by Type of Objective	18
Figure 2.7. Country Client Perceptions in FCV and Non-FCV Countries	20
Figure 2.8. Relationship between Quality at Entry and Project Preparation Time	22
Figure 2.9. IFC Investment Project Development Outcome Rating (annual data)	24
Figure 2.10. IFC Advisory Project Development Effectiveness Rating, Three-Year Moving Averages	26
Figure 2.11. IFC Investment Project Development Outcome Ratings by Industry Group	27
Figure 2.12. Factors Affecting IFC Investment Performance	30
Figure 2.13. MIGA Project Development Outcome Rating, Six-Year Rolling Basis	33
Figure 3.1. Steps in the Outcome Levels	37
Figure 3.2. Representative World Bank Project Objectives	39
Figure 3.3. Outcome Levels in IPF and DPF Projects	42
Figure 3.4. IFC Project and Market Claims' Outcome Levels	44
Figure 3.5. Representative Examples of IFC Claims	45
Figure 3.6. Ratings and Outcome Levels, by Instrument	49

Tables

Table 3.1. Ratings and Outcome Levels, by Instrument	48
Table 3.2. Ratings and Outcome Levels for Select Global Practices and Project Types	51
Table 3.3. Project M&E Rating and Evaluated Projects with Lack of Evidence, by Outcome Level	52



Abbreviations

COVID-19	coronavirus
CY	calendar year
DPF	development policy financing
FCV	fragility, conflict, and violence
FY	fiscal year
GP	Global Practice
IBRD	International Bank for Reconstruction and Development
ICRR	Implementation Completion and Results Report Review
IDA	International Development Association
IEG	Independent Evaluation Group
IFC	International Finance Corporation
IPF	investment project financing
M&E	monitoring and evaluation
MIGA	Multilateral Investment Guarantee Agency
MS+	moderately satisfactory or above
MTI	Macroeconomics, Trade, and Investment
<i>RAP</i>	<i>Results and Performance of the World Bank Group</i>
S+	satisfactory or better

All dollar amounts are US dollars unless otherwise indicated.



Acknowledgments

Rasmus Heltberg, task manager, led the work for this report under the supervision of **Alison Evans**, Director-General, Evaluation. The core team included **Mariana Branco**, **Claudia Figueroa Huidobro**, **Gaby Loibl**, **Xiaoxiao Peng**, **Stephen Porter**, **Melvin P. Vaz**, **Alena Lappo Voronetskaya**, and **Yi Yao**.

Other Independent Evaluation Group colleagues also made valuable contributions, including **Harsh Anuj**, **Ana Belen Barbeito**, **Leonardo Alfonso Bravo**, **Eric Cruikshank**, **Unurjargal Demberel**, **Hiroyuki Hatashima**, **Estelle Raimondo**, **Santiago Ramirez Rodriguez**, **Luis Alvaro Sanchez**, **Shiva Sharma**, and **Ichiro Toda**.

Maximillian Ashwill was the lead editor, and **JESS3** created the graphics.

The Independent Evaluation Group's quality enhancement panel included **Alison Evans**, **Oscar Calvo-Gonzalez**, and **Andrew Stone**. The external advisory panel included **Dr. Jörg Faust**, director of the Deval German Institute for Development Evaluation; **Tamar Manuelyan Atinc**, retired World Bank staff; and **Hans M. Boehmer**, retired World Bank staff and adjunct faculty, Columbia University.



Overview

This *Results and Performance of the World Bank Group (RAP)* assesses the World Bank Group's performance by analyzing the achievement of projects and program objectives through ratings and by classifying project objectives according to their outcome levels.

This report examines performance and outcomes from different perspectives using evidence from the Bank Group's results measurement systems. Previous *RAPs* have relied on the project and country program ratings that these systems collect. However, this report breaks with tradition and analyzes the results measurement systems' larger evidence base beyond ratings to classify outcome levels for World Bank and International Finance Corporation (IFC) projects. It also reviews how results measurement systems for select corporate priorities add up results and derives implications for the Bank Group's outcome orientation. Shifting the focus beyond ratings was partially done in response to the Board of Executive Directors' request for more evidence on the Bank Group's development outcomes and outcome orientation.

The data in this report cover a period ending in 2019 and do not show the coronavirus (COVID-19) pandemic's consequences for outcomes and performance, though the report identifies some implications for the Bank Group's COVID-19 response.


Part I: Assessing Performance through Ratings

World Bank Projects and Country Programs

Independent Evaluation Group project data for fiscal year (FY)19 show that 79 percent of World Bank lending operations were rated moderately satisfactory or above (MS+) at completion. This compares with 81 percent in FY18. Looking back over a longer period, the share of closed projects rated MS+ was 71 percent in FY09, declining to 63 percent in FY13 and rising since then. Measured by volume, 82 percent of lending operations were rated MS+ in FY19, staying relatively constant since FY13.

When results for FY12–14 and FY17–19 are compared, outcome ratings for investment project financing (IPF) show improved performance, from 68 percent MS+ to 81 percent MS+, and the share of development policy financing (DPF) operations rated MS+ decreased modestly from 72 to 69 percent.

Ratings increased in nearly all Regions and Global Practices. The Middle East and North Africa Region had the largest outcome ratings increases and now has the highest rating at 93 percent MS+ in FY17–19. Among the two Africa Regions, Western and Central Africa increased from 52 percent MS+ in FY12–14 to 71 percent MS+ in FY17–19. Eastern and



Southern Africa was also at 71 percent MS+ in FY17–19 and had remained stable at that level over the period. Project outcome ratings in countries affected by fragility, conflict, and violence (FCV) show improvement but continue to lag those in non-FCV countries. Between FY12–14 and FY17–19, the share of MS+ projects in FCV-affected countries increased from 69 to 77 percent compared with an increase from 69 to 81 percent in non-FCV-affected countries.

Other types of project ratings also increased over the past decade. Bank performance improved from 69 percent rated MS+ in FY13 to 84 percent in FY18 and 82 percent in FY19. Quality at entry ratings increased from 58 percent MS+ for projects that closed in FY14 to 75 percent for projects that closed in FY18 and FY19. Monitoring and evaluation quality ratings increased from 31 percent of projects rated substantial or above in FY09 to 51 percent rated the same in FY19. The improvement across these aspects of World Bank project performance, together with broadly conducive economic and institutional conditions in many larger countries during project implementation, helps explain the overall positive outcome ratings trends.

Beyond the project level, ratings for country program outcomes reached 72 percent MS+ in FY19, up from 51 percent MS+ in FY09. This increase occurred in International Bank for Reconstruction and Development countries, while country program outcome ratings stayed flat in International Development Association (IDA) and FCV-affected countries.


Country program performance was particularly low in FCV-affected countries because of external challenges, including large shocks for which country programs are often not sufficiently prepared. Weak political and technical capacity of governments in FCV-affected countries also explains the lower performance rating for projects focused on institutional and governance reform compared with those focused on service delivery.

Responding to COVID-19

Bank Group teams are preparing COVID-19 response projects under tight deadlines amid complex economic and public health contexts. Projects' quality at entry could suffer because the teams have less time and opportunity to conduct foundational work, client dialogues, and relationship building. Consequently, more frequent project and country program course corrections might be needed during implementation to respond to shocks and unforeseen circumstances and to mitigate issues associated with shorter project preparation time. Simpler procedures for restructuring and canceling projects could enable course corrections. Additionally, in low-capacity settings, teams could consider reducing country program scope when adding COVID-19 response components to avoid overtaxing countries' low implementation capacity.

IFC Investment and Advisory Projects

IFC investment project ratings for calendar year (CY)18 are the first to show a slight improvement after a 10-year decline. The CY18 data show that 43 percent of IFC investment



projects were rated mostly successful or better on development outcome, down from a peak of 75 percent in CY08 but slightly up from 40 percent in CY17. Measured by net commitment volumes and three-year moving averages, IFC's development outcome ratings declined from 83 percent rated mostly successful or better in CY07–09 to 43 percent in CY16–18 and 48 percent in CY17–19. Over this longer period, performance declined for all Regions, industry groups, country categories, and equity and loan instruments.

A combination of internal work quality issues, external risk factors, and broader market trends help explain IFC's performance trends. Issues with IFC staffing, incentives, accountability, and focus on volume targets over development results affected work quality. Market, country, and sponsor risks often distinguished higher-rated projects from lower-rated projects. Those with strong sponsors and business fundamentals coped better with market risks than projects without those characteristics. Additionally, projects that were better prepared to cope with currency devaluations and political and regulatory risks improved the likelihood of higher ratings. Broader market trends may have made IFC's business model more exposed to risk and weakened the pool of available projects with attractive risk-reward profiles. IFC has taken steps to improve work quality, focus on development results, grow the pool of bankable investment projects, and better identify risks and market opportunities.

Development effectiveness ratings began to improve for IFC advisory services projects evaluated in FY17–19, when 50 percent of them were mostly successful or better. The share rated mostly successful or better declined from 65 percent in FY12–14 to 38 percent in FY15–17. Measured by funding amounts, development effectiveness ratings declined from 70 percent mostly successful or better in FY12–14 to 33 percent in FY15–17 but then increased to 49 percent in FY17–19. Successful advisory projects often had strong client commitment, flexible and proactive supervision, and robust project monitoring and evaluation.

Multilateral Investment Guarantee Agency Projects

Multilateral Investment Guarantee Agency (MIGA) projects' development outcome ratings have continued on an increasing trend. These ratings increased from 64 percent satisfactory or better (S+) in FY07–12 to 69 percent S+ in FY13–18 when calculated by number of projects, and from 61 percent to 75 percent S+ when calculated by gross issuance amounts. MIGA projects in IDA and FCV-affected countries achieved high ratings—for example, 77 percent of MIGA projects in IDA countries were S+ compared with 63 percent in non-IDA countries in FY13–18. An analysis of MIGA projects in IDA countries found that MIGA promoted private sector investment by deterring political risks and resolving issues such as arrears payments by governments, for example.

Part II: Assessing Outcome Levels

Classification Framework

This *RAP* uses a theory of change framework to classify outcome levels, thus providing new information on the most common types of Bank Group project outcomes. The framework captures the intended and achieved outcomes of World Bank projects and the intended outcomes of IFC projects. The four outcome levels are the following:

- 1 · Outputs from Bank Group projects and activities
- 2 · Early outcomes such as a new capacity or better access to public services
- 3 · Intermediate outcomes such as a meaningful change in policy outcomes or beneficiaries' lives
- 4 · Long-term outcomes with systemic effects nationally or across sectors that contribute to general well-being

Outcome Levels

Project objectives cluster in clear outcome patterns depending on the sector and lending instrument. The patterns show that most IPF objectives focus on quality and access to services and cluster at level 2. However, IPF objectives in a few sectors (most notably agriculture and environment) have a clearer focus on end beneficiaries and cluster at level 3. Most DPFs, which focus on policy reform objectives outcomes, cluster at level 3, and recently approved IFC projects, which often focus on market creation objectives, cluster at level 3.

The relationship between projects' outcome levels and their performance rating is only modest and becomes insignificant when controlling for other factors. Ratings for projects with level 3 and 4 outcomes are modestly lower than for projects with level 2 outcomes, but the difference in ratings is insignificant when controlling for instrument and monitoring and evaluation quality. Many projects with higher-level objectives manage to achieve good Independent Evaluation Group ratings, in part by having strong results frameworks to measure outcome achievement. This finding suggests that there is no systematic trade-off between projects' outcome level and ratings, though it would not be realistic or desirable to expect all World Bank projects to have objectives at outcome level 3 or 4.

Differences in rating performance between IPFs and DPFs and between the lowest-rated Global Practice and other Global Practices appear more closely associated with levels of risk and the inherent difficulty in achieving policy and institutional reforms compared with service delivery improvements. Evaluation methods differ in reality between IPFs and DPFs, which may also play a role.



Thematic Area Outcomes

This *RAP* finds that the Bank Group clearly articulates higher-level outcomes for its global and thematic work in key thematic areas such as FCV, gender, and climate change. Results measurement systems in these thematic areas serve an essential accountability function by assuring that business units meet output and process targets, which are under the Bank Group's direct control. Yet a strong focus on monitoring targets can cause a risk-averse corporate culture and lead to box-checking behavior, meaning perfunctory rather than substantive compliance. Overall, systems that measure thematic area results do little to orient the Bank Group toward achieving higher-level outcomes.

Conclusions: Getting to Outcomes

This *RAP* concludes that the Bank Group often has limited evidence of its higher-level outcomes and can improve how its incentives and results measurement systems support outcome orientation. Projects' objectives need to balance realism and ambition, and therefore, one should not expect all projects to have higher outcome levels. There are more opportunities to gather evidence on broader outcomes at country program level.

Confronting trade-offs related to the purposes of the Bank Group's results measurement systems is necessary for improving outcome orientation. The Bank Group's results measurement systems collect evidence needed for ratings and for process and compliance monitoring. Systems collect little evidence on the Bank Group's contributions to higher-level outcomes, partly because such outcomes are hard to monitor and combine. At the project level, setting objectives and assessing achievements that can be attributed to Bank Group support continue to be important for the institution's accountability. Beyond the project level, there is a need to rethink the approach to collecting outcome evidence. A suitable approach would downplay ratings-based accountability, focus on contribution rather than attribution, and help stakeholders understand how different projects and types of Bank Group engagements collectively contribute to country-level outcomes over a longer period.



1 . Introduction

This is the 10th *Results and Performance of the World Bank Group (RAP)* by the Independent Evaluation Group (IEG). The *RAP* assesses the World Bank Group's performance by analyzing the achievement of project and program objectives through validated ratings and by classifying these objectives according to their outcome levels. It also explains key results and performance trends and discusses ways in which the Bank Group can continue to enhance its results measurement systems and outcome orientation.

Shifting the focus beyond ratings was partially in response to the Board of Executive Directors' request for more evidence on development outcomes and outcome orientation. It was also prompted by the recent capital increases to the International Bank for Reconstruction and Development (IBRD) and the International Finance Corporation (IFC), the International Development Association (IDA) Replenishment, and the need to report on a wider range of project and country outcomes from that expanded resource base.



Box 1.1. Key Terms in This Report

Project development objectives: World Bank projects' stated objectives framed as a positive outcome. In the International Finance Corporation's new Anticipated Impact Monitoring and Measurement system, *project claims* and *market claims* are similar statements of objectives or intended outcomes.

Outcome orientation: a term used when the World Bank Group generates credible evidence on the outcomes from its development interventions and uses this evidence to engage clients and adapt interventions and portfolios to bolster performance

Outcomes: changes in behaviors, conditions, or situations resulting from Bank Group activities. Outcomes include intended, unintended, positive, and negative changes.

Ratings: a measure of projects' and programs' success relative to objectives stated at approval or revised subsequently. Different aspects of projects and programs have separate ratings. For World Bank projects, the outcome rating measures how effectively and efficiently the project achieved its relevant objective.

Results: an all-encompassing term that refers to the outputs and outcomes from a development intervention

Results measurement systems: measurement systems that add up ratings and indicators from multiple projects and programs. The Bank Group has different primary results measurement systems for its main business lines, including World Bank, International Finance Corporation, and Multilateral Investment Guarantee Agency projects and country programs. The Bank Group also has different aggregated results measurement systems, such as the Corporate Scorecards and the results measurement systems for International Development Association, gender, and climate change.

Self-evaluation: the formal, empirical assessment of a project, program, or policy written by or for those in charge of the activity

Validation: the Independent Evaluation Group's independent, critical review of the evidence, results, and assessments from self-evaluations

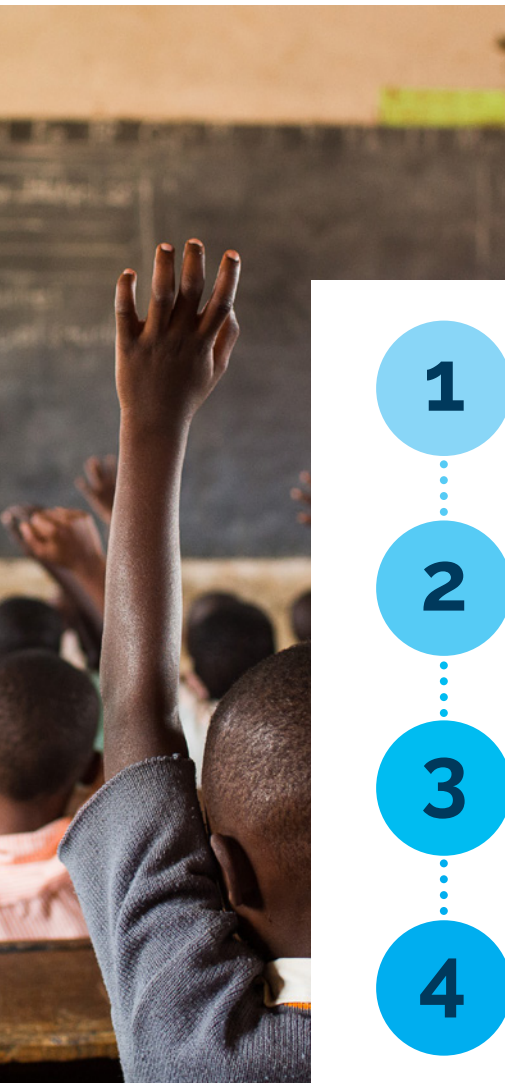
Source: Independent Evaluation Group.

This report examines ratings and outcomes from different perspectives using evidence from the Bank Group’s results measurement systems (see box 1.1 for key terms). Previous *RAPs* have relied on the project and country program ratings that these systems collect to understand the Bank Group’s results and performance. However, these results measurement systems contain much more evidence and many more indicators beyond project and program ratings.

This report breaks with tradition and analyzes this larger evidence base to also describe outcomes and classify outcome levels, particularly for closed and rated World Bank projects and for recently approved World Bank and IFC projects. It also reviews how results measurement systems for select corporate priorities add up results. To do so, the report synthesized sectoral theories of change derived from World Bank and IFC projects, among other sources, to build an outcome classification framework that could classify interventions’

stated objectives along a change pathway. Figure 1.1 defines this framework. In doing so, the report could examine different types and levels of outcomes and how these relate to performance, and assess the line of sight—or connection—between the Bank Group’s results measurement systems and higher-level outcomes.

Figure 1.1. Classification of Outcome Levels



1

LEVEL 1 Outputs

Activities and delivered outputs, such as knowledge products, goods, equipment, and services

2

LEVEL 2 Early or Immediate Outcomes

New capacities and better access to public, private, or environmental services

3

LEVEL 3 Intermediate Outcomes

Meaningful change in policy outcomes or beneficiaries’ lives

4

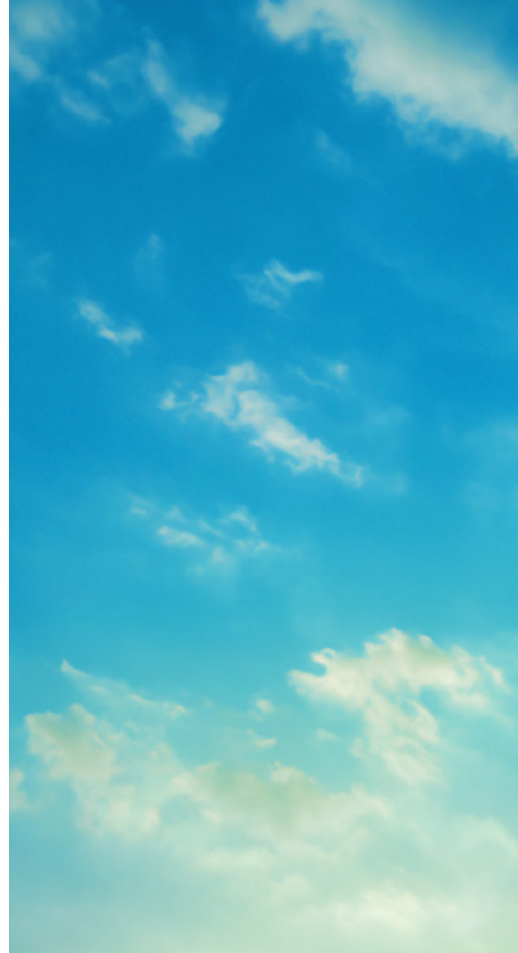
LEVEL 4 Long-Term Outcomes

Sustained long-term outcomes that eventually arise with sustained changes in delivery, governance, or citizens’ well-being

Source: Independent Evaluation Group. Refer to the full methodology in part II.

This report is in two parts. Part I is on performance as assessed through ratings and reports ratings trends for projects and country programs and identifies explanatory factors behind portfolio performance.

Part II is on assessing outcome levels and classifies objectives according to their outcome levels, examines links between performance and outcome levels, and discusses results measurement systems' outcome orientation. The *RAP* concludes with some key findings and implications for the Bank Group's coronavirus (COVID-19) pandemic response and its outcome orientation.



2. Part I

Assessing Performance through Ratings



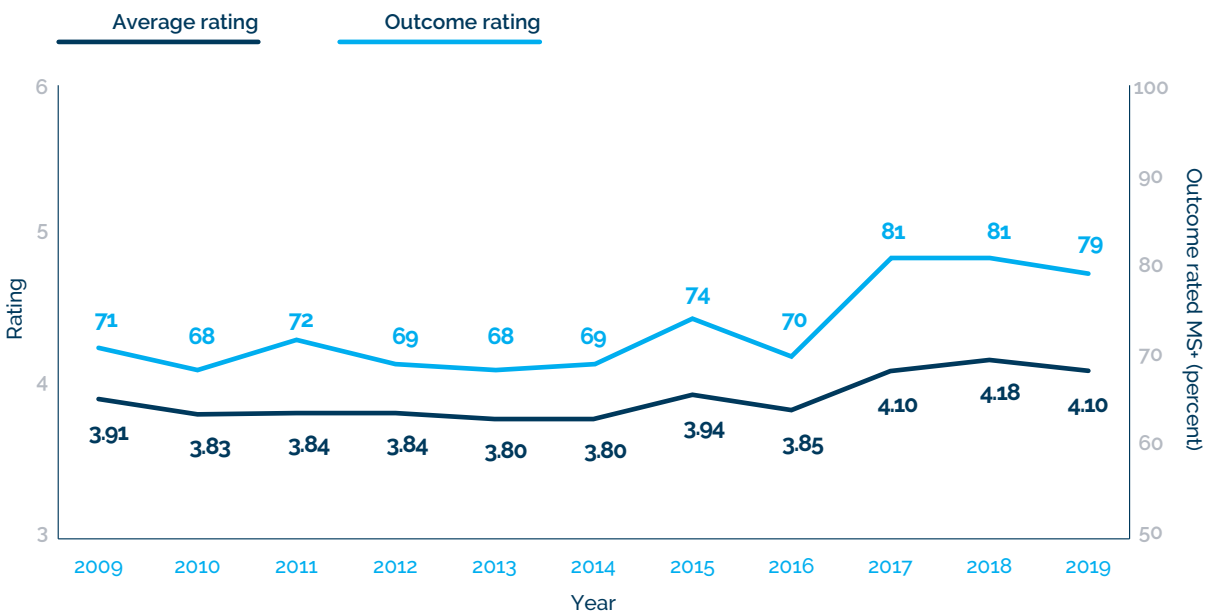
This chapter reports Bank Group ratings trends for World Bank projects, the Bank Group's country programs, IFC investment projects and advisory services, and Multilateral Investment Guarantee Agency (MIGA) projects. It also explains some major trends and patterns in ratings, focusing on World Bank projects and IBRD country programs' positive performance trends; programs in countries affected by fragility, conflict, and violence (FCV); and IFC's less positive performance. In line with common practice, the chapter treats ratings as a success metric. Ratings measure projects' achievement relative to objectives and targets stated at approval or revised subsequently. Ratings are not comparable across the three Bank Group institutions because of differences in mandates and business models.

World Bank Projects

Overall outcome ratings for World Bank lending are high. Of the 167 Project Completion Reports for projects that closed in fiscal year (FY)19 and were validated by IEG, 79 percent were rated moderately satisfactory or above (MS+) on achieving their stated outcomes. This is a slight decrease from 81 percent in FY18.

Looking back over a 10-year period, outcome ratings declined from 71 percent MS+ for project closures in FY09 to 68 percent MS+ in FY13, and they increased again to 81 percent MS+ in FY18 and 79 percent in FY19. A numerical conversion of the ratings scale, done to test the trend's robustness, shows the same pattern of ratings declines until FY13 and increases afterward (figure 2.1). Because of the improved project performance, outcome ratings increasingly cluster in the moderately satisfactory or satisfactory points of the scale.

Figure 2.1. World Bank Project Outcome Ratings, Annual



Source: Independent Evaluation Group.

Note | The dark blue line shows the numerical value of the six-point rating scale, which assigns 1 for highly unsatisfactory, 2 for unsatisfactory, and so on, with 6 being highly satisfactory. The light blue line represents the conventional percentage of projects rated moderately satisfactory or above. MS+ = moderately satisfactory or above.

Outcome ratings have been more stable over time when measured by project volume, increasing from 78 percent MS+ in FY08 to 82 percent in FY13, 84 percent in FY18, and 82 percent in FY19. IEG ratings for Bank performance improved from 69 percent of projects rated MS+ in FY13 to 84 percent in FY18 and 82 percent in FY19, reflecting better ratings for quality of supervision and quality at entry, the two components of the Bank performance rating (box 2.1).

Box 2.1. Aspects of Bank Performance

Quality at entry refers to the extent to which the World Bank identified, prepared, and appraised the operation so that it was most likely to achieve planned development outcomes.

Quality of supervision refers to the extent to which the World Bank identified and resolved threats to the achievement of development outcomes and to fiduciary aspects. The rating for quality at entry combined with the rating for quality of supervision determines the Bank performance rating.

Monitoring and evaluation (M&E) quality refers to the design and implementation of the project's M&E arrangements and the extent to which the data are used to improve performance. M&E quality is not a formal dimension of the Bank performance rating, though aspects of M&E overlap with quality at entry and quality of supervision.

Source: Independent Evaluation Group.

Outcome ratings increased in most parts of the portfolio. To have robust sample sizes, IEG compared the project closings of three-year cohorts. It compared FY12–14, when outcome ratings were at their lowest point (69 percent MS+), with FY17–19, when ratings were 80 percent MS+ (figure 2.2). Ratings for investment project financing (IPF) operations rose from 68 percent MS+ to 81 percent, and ratings for development policy financing (DPF) operations declined from 72 percent MS+ to 69 percent between FY12–14 and FY17–19, based on preliminary FY19 data.

Outcome ratings moved upward in nearly all Global Practices (GPs). Outcome ratings decreased in the Macroeconomics, Trade, and Investment (MTI) GP, which has the lowest ratings among all GPs, at 55 percent MS+ in FY17–19.¹ The MTI GP leads on many DPFs. Among the GPs with sizeable portfolios, the Education and Environment GPs' project ratings increased the most. Currently, the Education GP has the highest ratings, at 92 percent MS+. Part II examines the reasons for the ratings differential between IPFs and DPFs and between the highest- and lowest-rated GPs.

Ratings for projects in IBRD countries increased from 71 percent MS+ in FY12–14 to 82 percent in FY17–19. Ratings for projects in IDA countries increased from 68 percent MS+ to 78 percent over the same period. Ratings increased in nearly all Regions. The Middle East and North Africa Region saw the largest outcome ratings increases and now has the highest rating, at 93 percent MS+ in FY17–19. The Africa Region was split into two vice presidential units,

¹ Again, the fiscal year (FY)19 data is preliminary, so these numbers will change as more projects complete their evaluations.

effective July 1, 2020. Although the two Africa Regions were both at 71 percent MS+ in FY17–19, their trends differ. Western and Central Africa increased from 52 percent MS+ in FY12–14, but Eastern and Southern Africa remained stable from 72 percent MS+ in FY12–14.


Outcome ratings in FCV-affected countries increased modestly but remained below those in non-FCV-affected countries. Projects in FCV- and non-FCV-affected countries were both at 69 percent MS+ in FY12–14. Outcome ratings rose to 77 percent MS+ in FCV-affected countries in FY17–19 compared with 81 percent in non-FCV-affected countries (figure 2.2).

Figure 2.2. Project Outcome Ratings, FY12–14 and FY17–19 (percent rated MS+)



Source: Independent Evaluation Group.

Note | DPF = development policy financing | FY = fiscal year | MS+ = moderately satisfactory or above
 | FCV = fragility, conflict, and violence | IPF = investment policy financing



The underlying improved performance trends are also seen in higher ratings for projects' quality at entry (see definition in box 2.1). The share of MS+ quality at entry ratings increased from 58 percent MS+ for projects that closed in FY12–14 to 75 percent for projects that closed in both FY18 and FY19. Quality at entry ratings increased in all Regions and all Practice Groups except for Equitable Growth, Finance, and Institutions. Projects in FCV-affected countries have similar quality at entry ratings: 73 percent MS+ in FY18, a pattern also seen in previous years.

The World Bank maintained strong quality at entry even as it responded to the global financial crisis and increased its annual commitments to client countries by 130 percent. In fact, quality at entry improved for projects that had been approved since FY09. In part, this was possible because the World Bank increased the size of projects under preparation during the global financial crisis more than it increased the number of new projects.²

Monitoring and evaluation (M&E) quality ratings increased over the past 10 years for projects in all Practice Groups and Regions. M&E ratings rose from 31 percent of projects rated substantial or above in FY09 to 51 percent rated the same in FY19. All Regions increased M&E ratings over this period, and the Middle East and North Africa Region's ratings increased the most.

The overall increase in M&E quality ratings masks variation among GPs. Only four GPs achieved good M&E ratings substantial or above on at least half of their projects in FY16–18: Social Protection and Jobs (72 percent); Education (64 percent); Health, Nutrition, and Population (55 percent); and Urban, Resilience, and Land (51 percent). There have been many efforts to enhance tools, guidance, and training for staff to strengthen project M&E quality. Some examples include focusing attention on theories of change in project documents, restructuring projects to improve results frameworks, and building staff capacity by training existing staff and recruiting dedicated M&E specialists. Even so, interviews and desk reviews suggest that project M&E struggles for attention amid competing operational agendas.

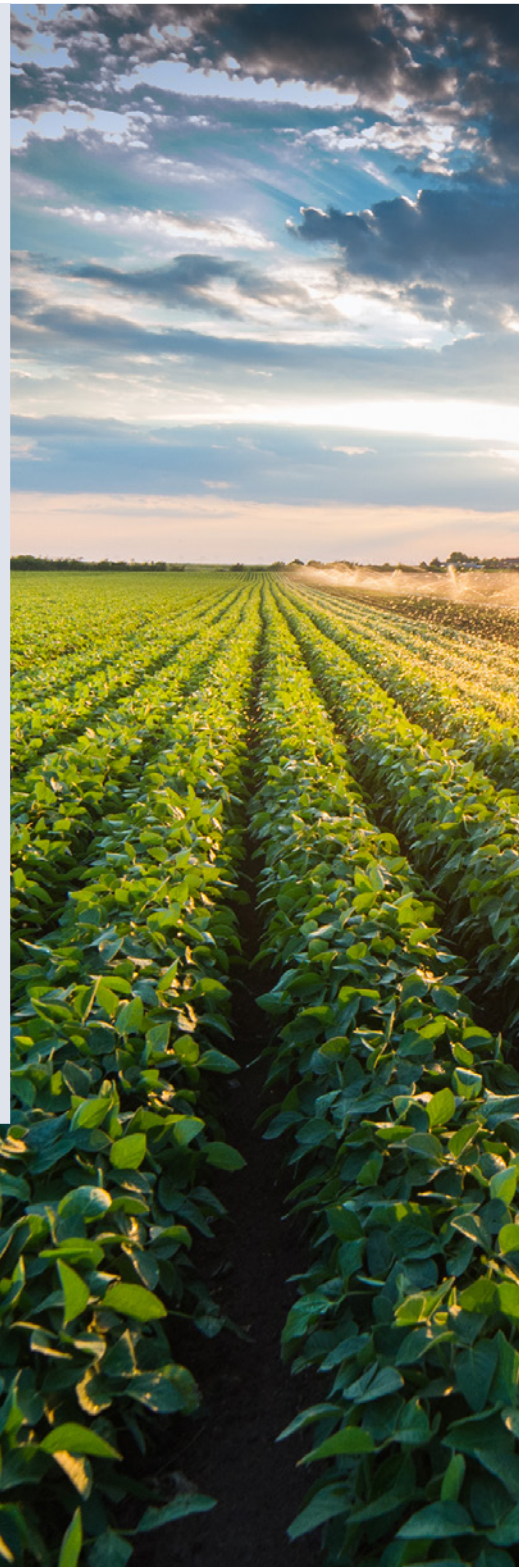
About 60 percent of the projects that closed between FY07 and FY18 have a mismatch, or disconnect, between the rating given to M&E quality in the last supervision report (the Implementation Status and Results Report) and IEG's validation of M&E quality based on the Implementation Completion and Results Report. The size of the mismatch varies quite widely by GP. It is possible that optimism bias is affecting assessments of M&E quality during implementation. The elements of what drives M&E quality are rather intuitive, as described in box 2.2.

² The World Bank nearly doubled the average size of new projects, from \$87 million in FY05–07 to \$157 million in FY09–10 (see also World Bank 2012).

Box 2.2. Elements of Monitoring and Evaluation Quality

Good project monitoring involves collecting the right data and using it in the right way. Projects with successful monitoring and evaluation (M&E) have outcome indicators that reflect project objectives without being too complicated. These projects plan and execute data collection that is computerized, quality controlled, aligned with client systems, and integrated into the operation rather than an ad hoc process. Teams use the data to track progress and identify implementation challenges. For example, an irrigation project in Mozambique had a specific project objective with clear, measurable, and directly linked indicators. The theory of change was sound, data collection was planned and executed regularly, weaknesses were corrected, and the team used the data to track progress, adjust the results framework during restructuring, and document project outcomes. Even better M&E also ensures country ownership over M&E arrangements, seeks to embed project M&E into client monitoring systems, and focuses on collecting useful data that can inform project implementation (versus more compliance-focused data). By contrast, projects with unsuccessful M&E had overambitious or complicated data collection plans and unclear results frameworks, resulting in delayed baseline data, irregular reporting, and information that lacked credibility.

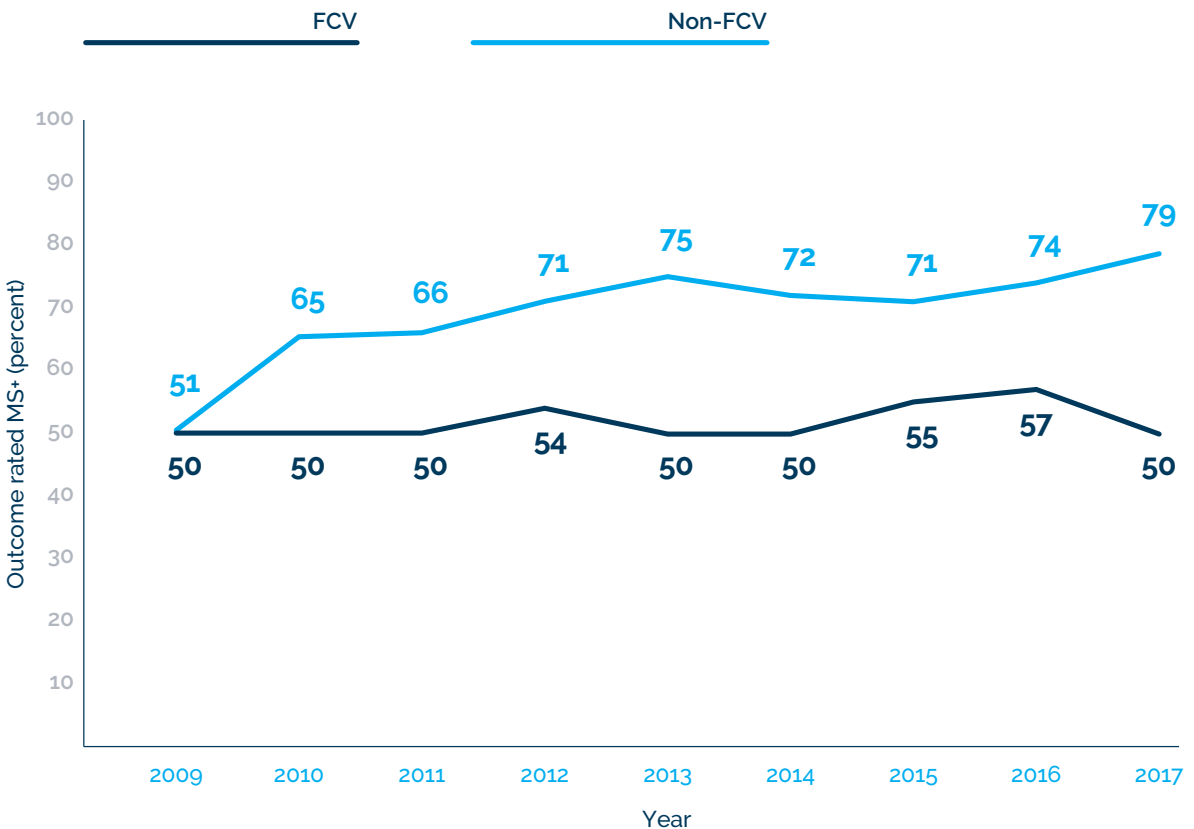
Sources: Independent Evaluation Group; World Bank 2016a.



Country Programs

Bank Group country program outcome ratings have improved over the past 10 years in IBRD countries but not in IDA and FCV-affected countries. Bank Group country program outcome ratings increased from 51 percent MS+ in FY09 to 74 percent in FY17 across all reviewed country program cycles. However, country program outcome ratings stayed flat in IDA and FCV-affected countries. These data are after smoothing, as explained in box 2.3. Among the six Regions, Europe and Central Asia and South Asia had the highest country program ratings in FY08–19, both at 79 percent MS+, and Africa, and East Asia and Pacific had the lowest, at 44 and 57 percent MS+, respectively. FCV-affected countries had lower country program outcome ratings over the period, at 50 percent MS+, compared with 66 percent MS+ for non-FCV (figure 2.3).

Figure 2.3. Country Program Outcome Ratings, FCV and Non-FCV Countries



Source: Independent Evaluation Group.

Note | The dark blue line shows the numerical value of the six-point rating scale, which assigns 1 for highly unsatisfactory, 2 for unsatisfactory, and so on, with 6 being highly satisfactory. The light blue line represents the conventional percentage of projects rated moderately satisfactory or above. MS+ = moderately satisfactory or above.

Box 2.3. Smoothing Country Program Ratings

This *Results and Performance of the World Bank Group* uses a new data smoothing method to compare project ratings across country programs. The Independent Evaluation Group conducts reviews of Completion and Learning Reviews (CLRs) for country programs at the end of every country program cycle, usually every four to five years. With only about 20 CLR reviews per year, the sample size is too small to allow many comparisons and identify meaningful trends. To overcome this data challenge, this report smooths annual data fluctuations by averaging country program outcome ratings over the four-to-five-year CLR period versus just the CLR's exit year. This method increases the number of data points per year and smooths country program outcome ratings over time.

Source: Independent Evaluation Group.

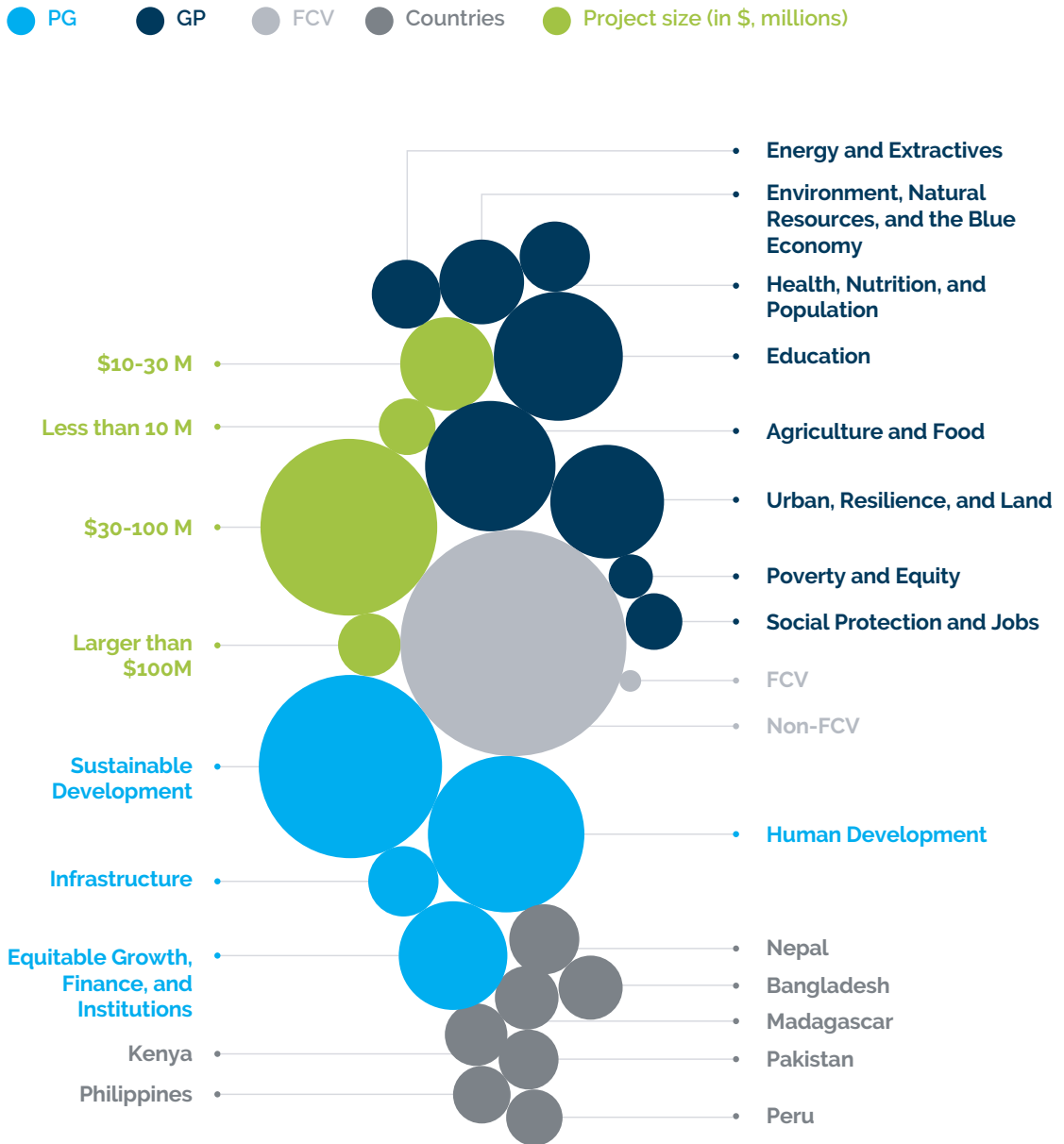
Explaining the Trends

The World Bank operates within country programs. Transforming its technical and financial support into results depends on both the country's capacity and economic environment and the quality of the World Bank's support. This *RAP* explores some of these external and internal factors further for World Bank projects and Bank Group country programs. It finds that improvements in project design, M&E, and supervision, combined with broadly conducive economic and institutional conditions during project implementation (that is, before the pandemic) in many of the larger countries, help explain the overall positive ratings trends. The worse performance in FCV-affected countries can partly be explained by difficult context and large shocks for which country programs in those countries were not sufficiently prepared.

IEG used decomposition analysis to account for the factors behind the increase in project outcome ratings between FY12–14 and FY16–18. The analysis decomposed the overall increase in World Bank project ratings over the period into changes in the size of different portfolio elements (such as Region, country, GP, lending instrument, and so on) and changes in the ratings for the portfolio elements. Figure 2.4 shows how much each portfolio element contributed to the total ratings increase. Decomposed this way, the increased portfolio share of projects in the South Asia Region (from 10 to 15 percent of the total portfolio size), together with modest improvements in project ratings, was an important contributor to improved performance ratings overall. Bangladesh, China, and Pakistan all had growing ratings and portfolios, thus increasing the total. IPF projects were the biggest contributor to improved average project outcome ratings.



Figure 2.4. Decomposing the World Bank Project Rating Increase over FY12–14 and FY16–18



Source: Independent Evaluation Group.

Note | The circle sizes represent how much each portfolio element contributed to the total ratings increase.

FCV = fragility, conflict, and violence	GP = Global Practice	PG = Practice Group
FY = fiscal year	M = millions	



A conducive institutional and economic environment and good performance in many of the World Bank's larger client countries contributed to improved project ratings. Many of the larger client countries saw good rates of economic growth and an uptick in their Country Policy and Institutional Assessment scores over this period.³ Studies have found a positive and statistically significant influence of economic growth and Country Policy and Institutional Assessment score on World Bank projects' performance (Geli, Kraay, and Nobakht 2014; World Bank 2018b). IEG's qualitative analysis of 14 projects rated highly satisfactory and 14 rated highly unsatisfactory found that the successful projects often benefited from a conducive context with strong political support and an enabling policy and regulatory framework. The opposite was true for the unsuccessful projects, which also suffered from political instability and clients' weak implementation and coordination capacity.



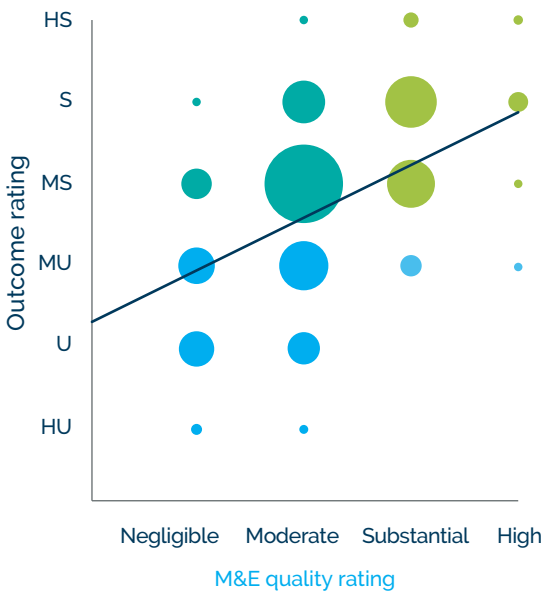
³ The Country Policy and Institutional Assessment score is an indicator of countries' policy framework and institutional capacity.

There is evidence of improvement across several aspects of the World Bank’s work quality. Most projects are designed well, as judged from the positive quality-at-entry ratings. The increase in projects’ M&E quality helps explain the increasing outcome ratings. Ratings methodology plays a role because IEG gives poor ratings to projects with insufficient evidence of their achievement. Regression analysis that attempts to control for the role of ratings methodology has shown that World Bank projects with good-quality M&E tend to have substantially—and statistically significant—higher ratings on outcomes than similar projects do (Raimondo 2016). The correlation between M&E quality and outcome ratings has held up over time and, in fact, has increased somewhat. So when outcome ratings are plotted against M&E quality, the slope has become modestly steeper (figure 2.5). The analysis of 14 projects rated highly satisfactory and 14 rated highly unsatisfactory found that M&E data collection and use of data for decision-making was one of the most frequent distinguishing factors. IEG ratings for supervision quality are also high, at 86 percent MS+ in FY19. This matters because studies have found that the task team’s ability to identify and mitigate potential risks to the project during supervision improves project outcome ratings.

Figure 2.5. Outcome Rating Plotted against M&E Quality Rating

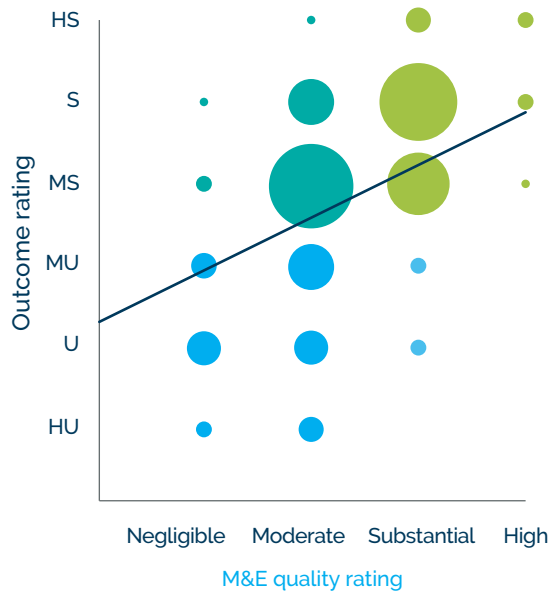
a. FY09–11

- Both high
- High M&E vs. low outcome
- Both moderate
- Both low



b. FY16–18

- Both high
- High M&E vs. low outcome
- Both moderate
- Both low



Source: Independent Evaluation Group.

Note | Circle sizes indicate how many projects fall in each category.

HS = highly satisfactory

S = satisfactory

MS = moderately satisfactory

MU = moderately unsatisfactory

U = unsatisfactory

HU = highly unsatisfactory

FY = fiscal year

M&E = monitoring and evaluation

Project outcomes can be achieved despite serious challenges if the task team can identify risks early, elicit support from managers, and act quickly to mitigate these risks, for example, by restructuring the project.⁴ The analysis of projects rated highly satisfactory found that these projects often benefited from collaborative supervision (active engagement of clients and partners, local presence, and a good mix of skills in the World Bank team) and timely reactions to challenges. Some non-IEG data also point to World Bank performance often being strong in the field. Country Opinion Surveys since 2012 indicate that country clients generally perceive the Bank Group positively as a long-term partner that collaborates well with government and contributes quality knowledge work, especially on good development and M&E practices. Survey respondents in a different survey conducted by AidData perceived the World Bank to be among the most influential donors, with particularly high influence of its knowledge products (Custer and others 2015).⁵



The worse performance in FCV-affected countries can be explained by a vicious cycle that these countries face in which large shocks prevent them from building capacity and improving governance. This has to be understood in a context of a somewhat rigid results framework architecture that requires forecasting results and is not sufficiently adaptable to dynamic circumstances, shocks, and high levels of uncertainty. These factors affect country program outcomes in various ways.

In a sample of 15 FCV-affected countries, all experienced large shocks, such as Ebola outbreaks, disasters, oil price shocks, and political crises. These shocks altered national priorities, prevented countries from building stable and credible institutions, and compelled the country team to reallocate resources and adjust country programs' implementation. Political shocks and armed conflict (for example, in Madagascar and the Republic of Yemen) are especially challenging. The reduced staff presence during a political crisis naturally made it hard to reengage and achieve program objectives after the crisis subsided.

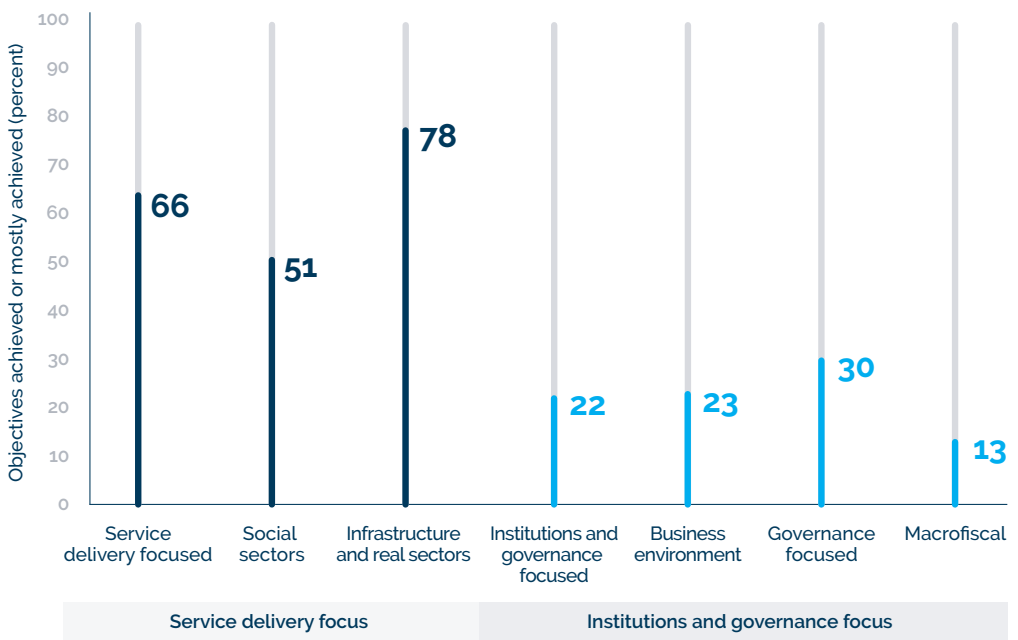
⁴ Two World Bank reports (2016a, 2018b) summarize the evidence, including an internal audit study.

⁵ Unlike the International Finance Corporation (IFC), the World Bank does not have a rating system for its knowledge products. Perception surveys suggest they can be influential. AidData's survey in Custer and others (2015) was updated in AidData's 2014 Reform Efforts Survey Aggregate Data Set (2017).



Institutional and governance reforms in FCV-affected countries are often unsuccessful. About half of country program objectives in FCV-affected countries focused on institutional and governance reforms, and the other half focused on infrastructure development and public service delivery. Only 22 percent of objectives that focused on institutional and governance reform were achieved or mostly achieved, compared with 66 percent of objectives focused on service delivery (figure 2.6).⁶ Institutional and governance reforms are harder to insulate from FCV-affected governments’

Figure 2.6. Ratings for Country Program Objectives by Type of Objective




Source: Independent Evaluation Group.

Note | MS+ = moderately satisfactory or above

weak political and technical capacity and need more time to achieve objectives than public service delivery projects do, which could explain the lower achievement rates of these reforms. The longer timeline exposes these reforms to more shocks and more government and World Bank staff turnover.

⁶ This is according to Completion and Learning Report Reviews, the Independent Evaluation Group’s (IEG) reviews of closed country programs.



Overburdened country programs performed worse in response to large shocks and crises. In this context, *overburdened* refers to FCV-affected country programs with weak relevance and selectivity. Bank Group country programs performed better during shocks when they limited and consolidated interventions, such as in Haiti, Kosovo, Lebanon, Nepal, and Timor-Leste. Some other shock-affected countries saw an influx of new projects or made existing projects more complex, overstressing the World Bank's and clients' capacity.

Most FCV-affected country programs were already stretched to capacity before the shocks occurred. In a sample of 15 recently evaluated FCV-affected country programs, 11 had low or weak relevance, defined as the likelihood a program will achieve its intended objectives given the program's resources and instruments. Nine programs had low or weak selectivity, which is defined as concentrating resources on priority objectives in a way that maximizes development impact and does not overburden the client's or the World Bank's implementation capacity. Eight programs were neither relevant nor selective. Four FCV-affected country programs were both relevant and selective, and three of these—Liberia, the Solomon Islands, and The Gambia—performed well despite major shocks.

Mirroring these findings, IEG's project validations often find that project designs that are too complex relative to clients' capacity lead to weak ratings in FCV-affected countries. Some econometric studies have associated active conflict, inflation, natural resource dependence, and distortionary trade, fiscal, and monetary policies with lower project performance.⁷

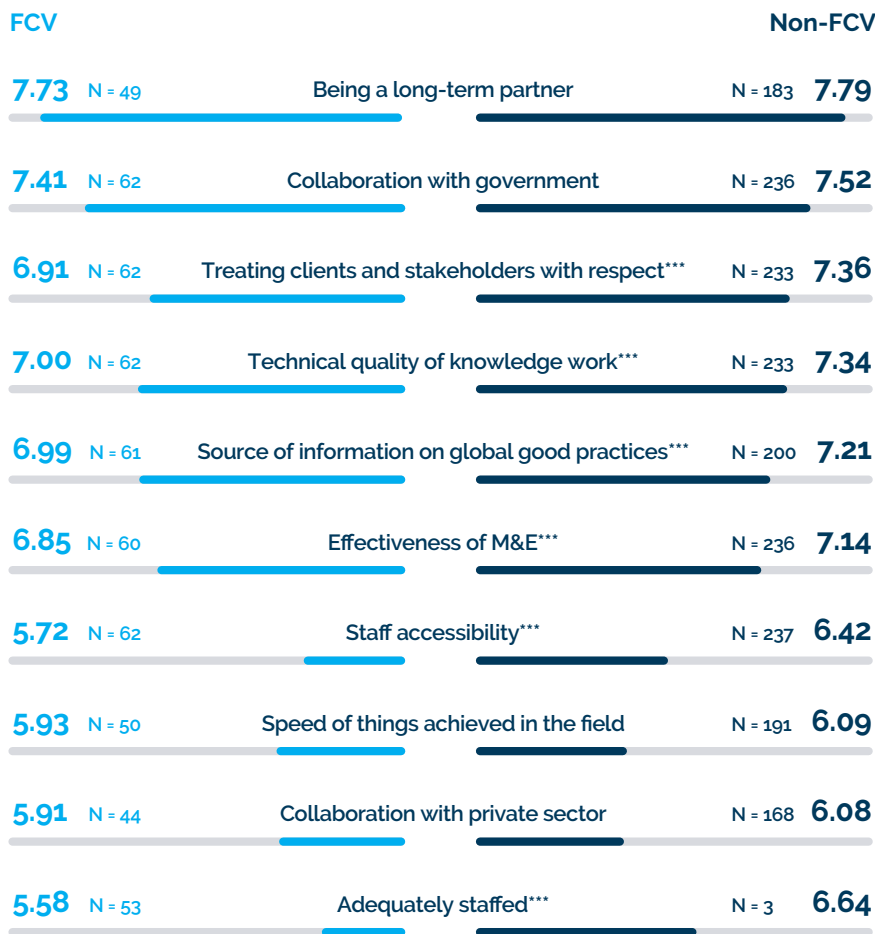
Staff quality and presence also matters for performance. Studies have linked the quality and stability of the project's task team leader to project performance—see, for example, Denizer, Kaufmann, and Kraay 2011; Geli, Kraay, and Nobakht 2014; Ralston 2014; Moll, Geli, and Saavedra 2015; and World Bank 2016b. Yet client perceptions of the Bank Group staff's availability and the quality of its work are often less favorable in FCV-affected countries. On most Country Opinion Survey questions, perceptions of the Bank Group are worse in FCV-affected countries than in non-FCV-affected countries. This is especially true of the Bank Group's respectful treatment of clients and stakeholders, the technical quality of its knowledge work, its value as an information source on global development practices, its project M&E, and its staff accessibility (figure 2.7). Similarly, responses in AidData's survey on the World Bank's perceived influence were markedly lower in FCV-affected countries than in others.⁸ It is not clear why FCV clients respond less favorably to perception surveys. The Bank Group has increased its budget and staff resources for FCV-affected countries over time,

⁷ World Bank 2018b summarizes the studies.

⁸ According to a custom calculation that AidData provided to the *Results and Performance of the World Bank Group*, the average score for World Bank influence was 3.76 among non-fragility, conflict, and violence-affected respondents, compared with 3.28 for respondents in countries classified as fragility, conflict, and violence-affected in FY14, the year before the survey. This is based on data documented in Custer and others (2015) and updated in AidData's 2014 Reform Efforts Survey Aggregate Data Set (2017).

though recruiting qualified staff to work in FCV-affected countries has often been difficult, something that the Bank Group’s strategy for FCV (2020–25) is aiming to address through enhanced support, training, and incentives for staff working in fragile settings.

Figure 2.7. Country Client Perceptions in FCV and Non-FCV Countries



Source: Independent Evaluation Group, based on World Bank Group Country Opinion Survey data collected annually from 2012 to 2019.

Note | All scores except for one are measured with the following Likert scale:
 1 = to no degree at all; 10 = to a very significant degree.
 Technical quality of knowledge work is measured with the following Likert scale:
 1 = very low technical quality; 10 = very high technical quality.
 Averages (“N”) are based on number of country-years. Statistical significance is for difference of means tests between question responses in FCV and non-FCV countries. Two-sample mean tests are used, assuming equal variances.
 | FCV = fragility, conflict, and violence | M&E = monitoring and evaluation | *** $p < .01$.

Responding to COVID-19 and Other Shocks

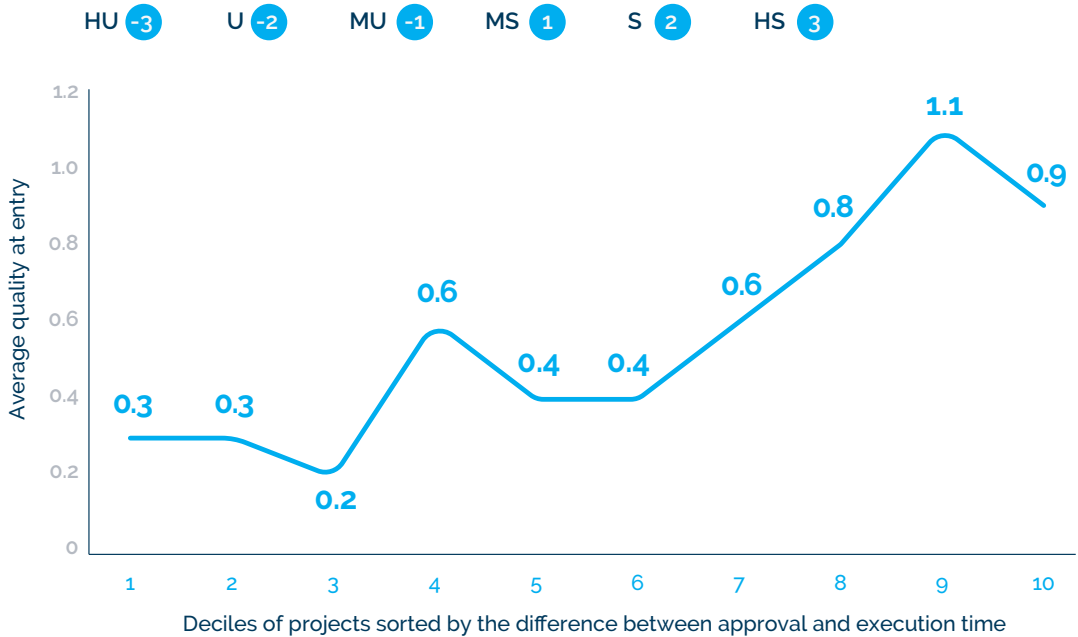
The study of shocks and their impact on project and program results can also contribute some insights for the World Bank’s ongoing pandemic response. Teams are preparing pandemic response projects (many of which are new rather than additional financing) under tight time pressures and amid complex political, economic, and public health contexts and logistical challenges, such as the inability to travel or conduct meetings in person. According to IEG evaluations, there is sometimes less time for data collection, technical studies, learning from past lessons, and designing strong results frameworks when the World Bank rushes to prepare crisis responses—see, for example, World Bank 2010a, 2010b, and 2017. World Bank (2019) analyzed comprehensively the factors that influence quality at entry and found that foundational work matters. Less foundational work limits the World Bank’s understanding of local policy, capacity, and institutions and its ability to fine-tune procurement arrangements and other elements of project design. The logistical challenges could adversely affect the World Bank’s local staff presence and ability to build trusting relationships and partnerships—factors that World Bank (2019) also found critical for quality at entry.⁹



⁹ Mirroring this, econometric research has linked project outcomes to the project team’s access to time, budget, and knowledge (see Ika 2015; and World Bank 2016b, 2017).

There is a statistical association between time pressures during preparation and projects' quality at entry. IEG calculated a variable for project preparation time in a sample of more than 3,000 evaluated projects.¹⁰ Projects in the first three deciles of this variable, meaning projects with low preparation time relative to duration, were rated significantly lower on quality at entry than projects at or above the median of project preparation time (figure 2.8).¹¹

Figure 2.8. Relationship between Quality at Entry and Project Preparation Time



Source: Independent Evaluation Group.

Note | HS = highly satisfactory | MU = moderately unsatisfactory
 | HU = highly unsatisfactory | S = satisfactory
 | MS = moderately satisfactory | U = unsatisfactory

¹⁰ This index is the absolute difference in months between projects' approval time (time from inception to approval) and projects' duration (time from effectiveness to project close).

¹¹ The score for quality at entry was calculated by converting the 6-point scale into numerical values:

highly unsatisfactory = -3	moderately satisfactory = 1
unsatisfactory = -2	satisfactory = 2
moderately unsatisfactory = -1	highly satisfactory = 3



Robust implementation support can counter shocks, problems, and quality at entry weaknesses. The difference between projects rated highly satisfactory and highly unsatisfactory was less about the presence of shocks or the number of supervision missions but instead about the World Bank teams’ timeliness in flagging concerns, taking corrective measures, complying with mandated safeguards, undertaking Mid-Term Reviews, revising objectives, and collecting data. These often distinguished successful and unsuccessful projects. For example, Somalia’s Emergency Drought Response and Recovery Project, which IEG rated highly satisfactory on outcomes, was prepared in five weeks and required complex support to implement. It involved intense collaboration and overcoming institutional differences between the World Bank and the International Committee for the Red Cross on rules and procedures for M&E, procurement, financial management, and even protocols for communicating with government officials.



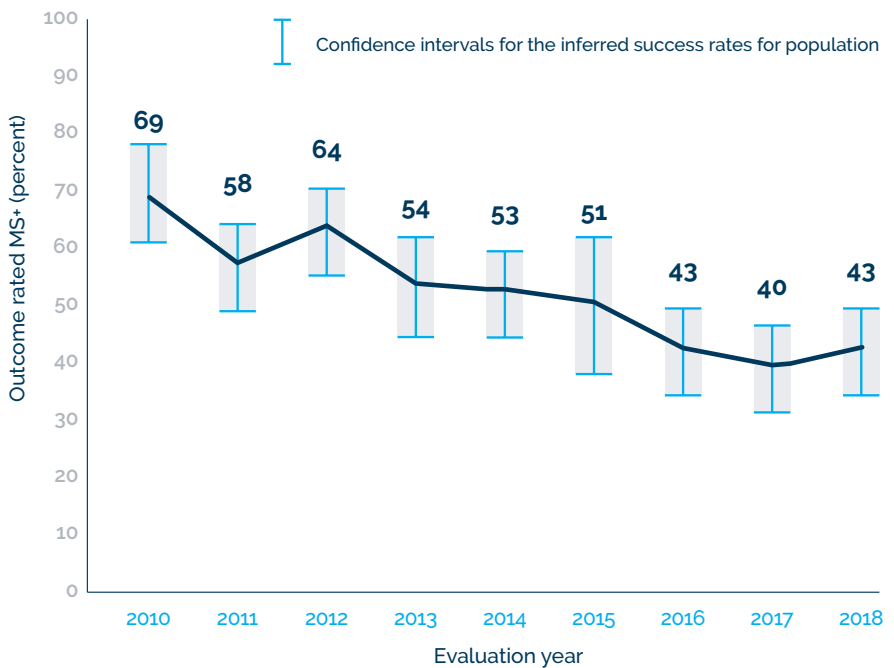
IFC Projects

Investments

IFC investment projects' development outcome ratings have declined over the past 10 years, but there are early signs that this decline has stopped or may be starting to reverse. IFC development outcome ratings declined from a peak of 75 percent of projects rated mostly successful or better by IEG in calendar year (CY)08 to 40 percent in CY17 and 43 percent in CY18 (figure 2.9). These ratings are based on a stratified random representative sample, which in CY18 covered 99 projects, or 39 percent of all projects approved in CY13 and eligible for evaluation. Average ratings can also be measured by net commitment volumes rather than the number of projects and by using three-year instead of annual averages. Calculated this way, IFC's development outcome ratings declined from 83 percent rated mostly successful or better in CY07–09 to 43 percent in CY16–18 and 48 percent in CY17–19. As these numbers suggest, the ratings decline may have stopped or reversed since CY17.¹²



Figure 2.9. IFC Investment Project Development Outcome Rating (annual data)



Source: Independent Evaluation Group.

Note | IFC = International Finance Corporation | MS+ = mostly successful or better.

¹² The tentative reversal in IFC's ratings trend is, however, within the margin of error, given that only a sample of IFC projects undergo ex post evaluation and that not all of the projects sampled for evaluation in the calendar year 2019 cohort have finished their evaluations.

IFC infrastructure projects' development outcome ratings fell from 63 percent mostly successful or better in CY13–15 to 40 percent in CY16–18. Development outcome ratings for projects involving oil and gas exploration and junior mining companies declined sharply (from 73 percent mostly successful or better in CY13–15 to 13 percent in CY16–18); IFC has halted or reoriented most of its oil, gas, and mining investing. Core infrastructure projects (that is, excluding oil, gas, and mining) were 46 percent mostly successful or better, which is similar to other Industry Groups. IEG's review of infrastructure projects contains another lesson with wider applicability for IFC project success. These projects have shown that essential client actions, such as obtaining operating permits or licenses or reporting monitoring data, must be completed before disbursing the equity investment to the client. This is because IFC is a minority shareholder with limited recourse or influence after investments are disbursed. These projects also show that IFC's early and continuous project engagement contributed to successful social and environmental ratings, particularly when companies expanded into different sectors and countries and thus benefited more from IFC's advice. In recent years, IFC has expanded its advice to existing and prospective clients on social, gender, environmental, and community engagement issues.

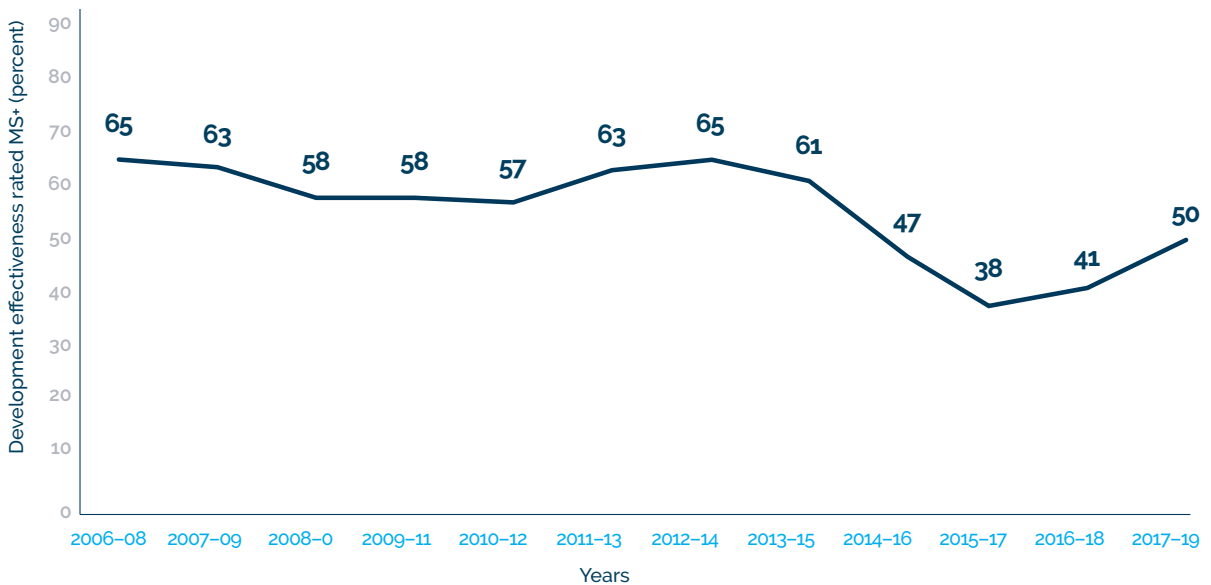




Advisory Services

Development effectiveness ratings for IFC advisory services projects show signs of improvement. Development effectiveness ratings peaked in FY12–14, when 65 percent of advisory projects were rated mostly successful or better (figure 2.10). This declined to 38 percent in FY15–17 before increasing to 41 percent for projects evaluated in FY16–18 and 50 percent for FY17–19 (based on very preliminary FY19 data and therefore subject to change).¹³ When calculated by the advisory project’s funding amount rather than the number of projects, development effectiveness ratings declined from 70 percent mostly successful or better in FY12–14 to 33 percent in FY15–17, before increasing to 49 percent in FY17–19.

Figure 2.10. IFC Advisory Project Development Effectiveness Rating
Three-Year Moving Averages



Source: Independent Evaluation Group.

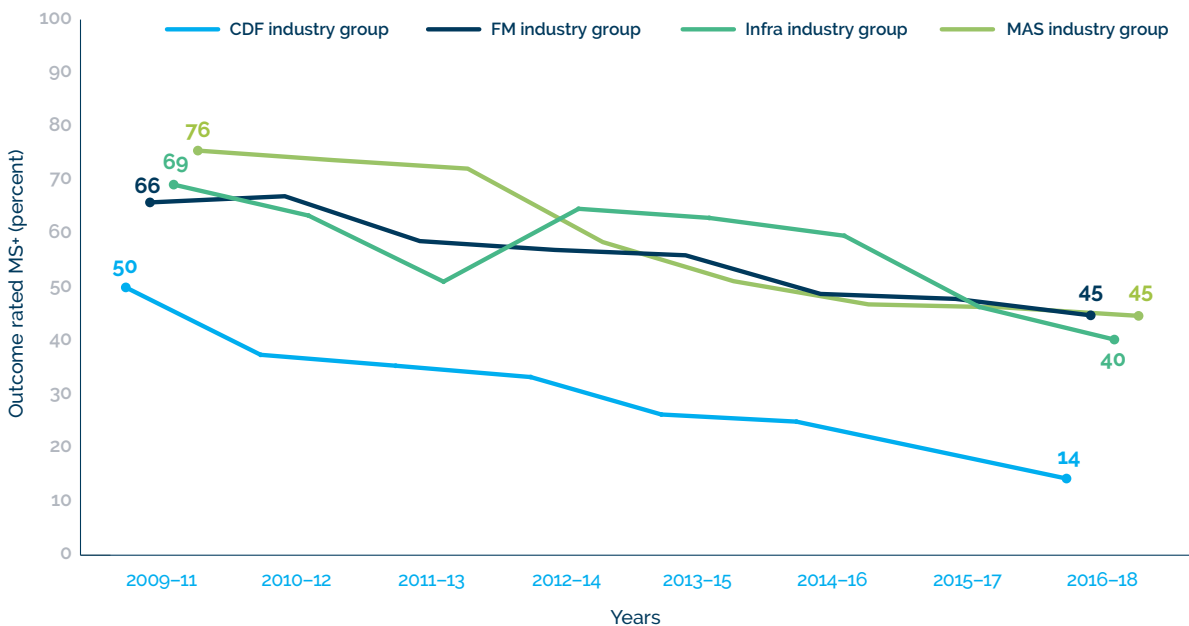
Note | IFC = International Finance Corporation | MS+ = mostly successful or better.

¹³ Although the FY17–19 estimate is based on 171 evaluated projects, the FY19 data are based on only 36 evaluated projects out of 54 projects sampled for evaluation. Estimates will therefore change as more projects finish their evaluations.

Explaining the IFC Trends


IEG researched many possible explanations for the long period of decline in IFC investments' development outcome ratings. The joint IFC-IEG underlying evaluation and ratings methodologies did not change during this period, so methods changes cannot explain IFC's ratings decline. The ratings trends for IFC investments differ substantially from those of the Asian Development Bank and the European Bank for Reconstruction and Development, which are the only other multilateral development banks with published ratings for private sector operations. Both institutions' development outcome ratings for private sector investment projects increased over the same 10-year period in which IFC's ratings dropped, so global economic conditions alone cannot explain the ratings decline. Additionally, IFC's development outcome ratings declined in all Regions; for all four industry groups (figure 2.11); in IDA-eligible, FCV-affected, and IBRD countries; for both equity and loan instruments; and in both greenfield and expansion projects. Therefore, major declines in specific project categories cannot explain IFC's ratings decline because declines were across the board. Furthermore, IFC's business volume stayed approximately the same over the past 10 years with no major investment increases in low-capacity countries, so rapid business growth cannot explain IFC's ratings decline. In fact, IFC ratings in IDA-eligible countries are slightly higher than in IBRD countries.

Figure 2.11. IFC Investment Project Development Outcome Ratings by Industry Group



Source: Independent Evaluation Group.

Note | CDF = Disruptive Technology and Funds | FM = Financial Markets | MAS = Manufacturing, Agribusiness, and Services.
 IFC = International Finance Corporation | Infra = Infrastructure
 MS+ = mostly successful or better



A combination of internal work quality issues, external risk factors, and broader market trends help explain IFC's investment performance trends. A joint IFC-IEG study from 2017 identified work quality and credit and country risks as significant drivers of investment projects' development outcome ratings. Staffing, incentives, organizational culture, focus on volume targets over development results, and diffused accountability were the main factors affecting IFC's work quality. IFC endorsed those findings and has since taken many steps to implement the joint study's recommendations, including setting up a vice presidential unit to focus on development results, seeking stronger country engagement with improved analytics, and screening projects ex ante for anticipated outcomes.¹⁴

External risk factors also influence projects' performance. IFC invests in many domestic, medium-size firms affected by a variety of risks. IEG's review of project validations found that market, country, and sponsor risks and transaction structuring were the factors most clearly associated with IFC investment projects' performance. IEG reviewed nearly two-thirds of the projects that it had evaluated from 2016 to 2018 and used a machine learning framework to analyze all 720 IFC investment projects that IEG evaluated between 2010 and 2018. Both reviews sought to identify factors associated with projects' success and underperformance, and both identified sponsor selection risks, market risks, country risks, and transaction structuring as the factors that most frequently distinguished projects with good ratings from less successful projects (figure 2.12). The machine learning algorithm clustered projects as high development outcome and high work quality (high work quality, 318 projects), and low development outcome and low work quality (low work quality, 213 projects).

- Sponsor risks (risks linked to the client company in which IFC invests) were 1.8 times more frequent in projects with low work quality than in those with high work quality. IFC knew the sponsor well in the positively rated projects (high work quality). Either the sponsor was a repeat client in good standing and had strong business fundamentals or IFC's due diligence had concluded that the client had the necessary knowledge and experience. By contrast, sponsors of projects with low work quality had started new business lines in which they lacked relevant experience or were highly leveraged.

¹⁴ IFC's managerial actions translate into ratings for the mature portfolio with a long delay. That is because the projects rated this year were approved years before the mentioned actions.



- Projects with high work quality coped better with market risks, which affected all types of projects. For example, slowdowns in visiting tourists and an oversupply of hotel rooms affected some tourism projects. Weak consumer demand caused by slowing economic growth and currency devaluations affected some agribusiness and forestry projects. Demand that was weaker than expected or competitors adding infrastructure capacity affected some infrastructure projects. Such market risks had only temporary effects on projects with sound underlying business fundamentals, strong sponsors, and enough liquidity. Market risks had the most lasting impacts on the success of projects without these strong fundamentals.
- Country risks increased in relative influence on investment projects. The most common country risks were currency devaluations and political and regulatory risks. These risks increased for projects with both low work quality and high work quality that IEG evaluated in 2013–15 and 2016–18. However, the machine learning algorithm found that in projects with high work quality, IFC and its clients adapted well to country risks. Sometimes, IFC successfully mitigated the impact of currency devaluations on projects through local currency loans. However, there are examples where IFC lent to clients in foreign currency even though local currency IFC loans were also available. Other projects with low work quality relied on anticipated regulatory changes for project viability. However, these changes often took much longer than anticipated or did not happen at all, adversely affecting project results.

- The quality of IFC’s transaction structuring, additionality, and sensitivity analyses varied between projects with high work quality and low work quality. Details vary across industry groups. Examples of strong IFC transaction structuring included good selection of IFC investment products, careful scrutiny of intragroup risks when investing in holding companies, rigorous analysis of market and exchange rate risks, and realistic consideration of a bank’s condition and priorities before investing in those banks.

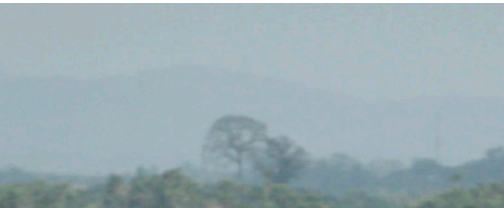
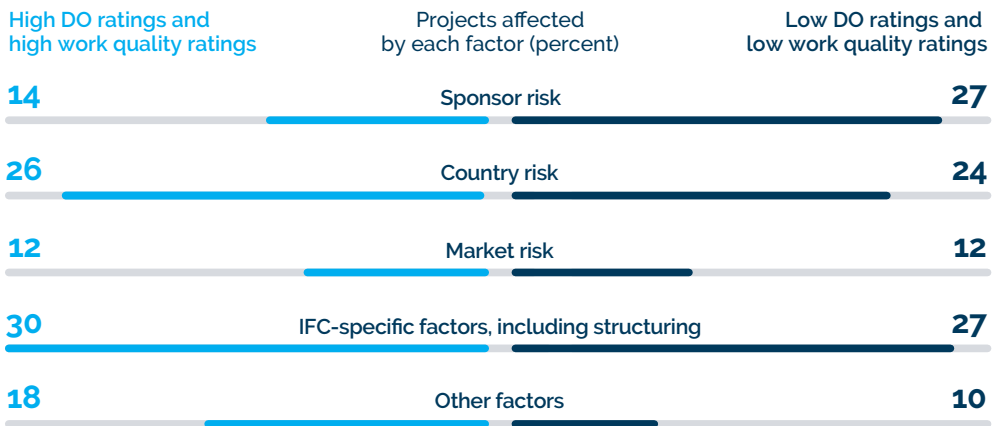



Figure 2.12. Factors Affecting IFC Investment Performance



Source: Independent Evaluation Group.

Note | DO = development outcome | IFC = International Finance Corporation





Broader market trends may have made IFC's business model more exposed to certain risks. IFC screens for risks when selecting projects, but there is a finite pool of repeat clients and bankable or viable investment projects, so IFC needs to accept certain risks when it invests. Moreover, the pool of viable projects available to IFC may have shrunk because rival financiers, both private and multilateral, have expanded into emerging and developing markets over the last few decades. A weaker pool of viable investment projects can translate into less attractive risk-reward profiles, thus contributing to the ratings decline. This means that better internal identification of risks during project preparation may not suffice. In recent years, IFC has taken many steps to grow the pool of bankable investment projects and to better identify market opportunities and constraints, as described in box 2.4. It has also taken other steps to increase its focus on outcomes, including providing specialist resources to advise teams, encouraging midcourse corrections, and introducing a new tool (Anticipated Impact Monitoring and Measurement) to assess and screen projects for expected outcomes prior to approval.¹⁵ Hopefully, these steps will help align IFC's business model to market, country, and sponsor risks, though it is too early to tell if they will improve ratings. Additional steps IFC could consider include enhanced tools and processes to identify and mitigate risks during supervision.


Box 2.4. IFC's Reforms to Strengthen Upstream Engagement

The International Finance Corporation (IFC) has prioritized upstream engagement in its strategy IFC 3.0. Upstream engagement can increase the number of bankable investment opportunities through regulatory reforms to unlock private investment and development of viable investment projects. To do so, IFC has updated its funding and operating model to encourage upstream engagement and invested significant resources in developing its project pipeline.

IFC has strengthened its focus on country outcomes through new IFC Country Strategies and analytical tools such as Country Private Sector Diagnostics and IFC Sector Deep Dives. These tools aim to provide a deeper understanding of market constraints and opportunities and help develop better coordinated upstream engagements with hopefully greater development outcomes. IFC has also integrated advisory teams into industry groups and introduced a new additionality framework.

Source: Independent Evaluation Group, based on documents from IFC.

¹⁵ See World Bank (2019, 19) for a fuller description of IFC's efforts to improve work quality.



For IFC advisory projects, project size and duration and a change of team leader had a statistically significant negative association with project success.¹⁶ Some of the larger and longer-lasting projects were riskier, for example, if they involved public sector clients and complex regulatory reforms such as those in business climate and public-private partnerships. Some of these larger advisory projects were more likely to encounter difficulties with political economy and counterparts' capacity compared with simpler projects with private sector clients. Such difficulties could increase in importance as IFC expands its upstream engagements and its program in challenging and fragile markets.

Other factors that mattered for IFC advisory projects' success included the client's commitment, IFC's flexible and proactive supervision, and robust project M&E.¹⁷ Client commitment was a major driver of advisory projects' success. Indications of client commitment include alignment with the client's established business plan or ongoing activities, client contributions to project costs, and level of seniority of interlocutor staff. Commitment can be fostered by aligning client and project objectives, involving clients closely in project design, and establishing a variety of client interlocutors beyond the project's day-to-day individual counterparts. The staff's patience and flexibility to respond to changing circumstances (such as government personnel changes) also contributed to success, and detecting signs of waning client commitment and restructuring projects accordingly proved important. However, such project restructurings were helpful only when clients showed continued commitment, for example, through responsiveness and engagement; otherwise, canceling the projects was preferable. IFC staff and managers' proactive involvement in decisions to restructure, cancel, or reduce the duration of projects was important because IFC consultants contracted to the project may lack the incentive to recommend such actions. Robust project M&E provides IFC teams with a more detailed understanding of projects' achievements and challenges so that they can adjust implementation as needed and achieve results. As reported in *RAP 2018*, IFC has worked to strengthen its work quality for some years, including through greater attention to projects' scope and results frameworks, self-evaluations, and staff training.

¹⁶ This is based on all 169 advisory projects evaluated between FY16 and FY18.

¹⁷ This is based on IEG's review of 42 advisory projects evaluated in FY18.

MIGA Projects

Ratings for MIGA projects' development outcomes increased over the past 10 years. Specifically, the ratings increased from 64 percent satisfactory or better (S+) in FY07–12 to 69 percent in FY13–18 (figure 2.13). When calculating ratings by gross issuance amounts, MIGA development outcome ratings increased from 61 percent S+ to 75 percent over the same time frame. These higher ratings have continued into the most recent ratings period. The increases are driven by higher ratings for MIGA projects in IDA countries (from 59 percent S+ in FY07–12 to 77 percent in FY13–18), in Europe and Central Asia (from 56 percent S+ to 73 percent), and in the Energy and Extractive Industries sector (from 67 percent S+ to 79 percent).

MIGA projects are rated at 77 percent S+ in IDA countries, which are a strategic focus for MIGA, compared with 63 percent S+ in non-IDA countries. Projects in FCV-affected countries also had high ratings at 88 percent S+ in FY13–18

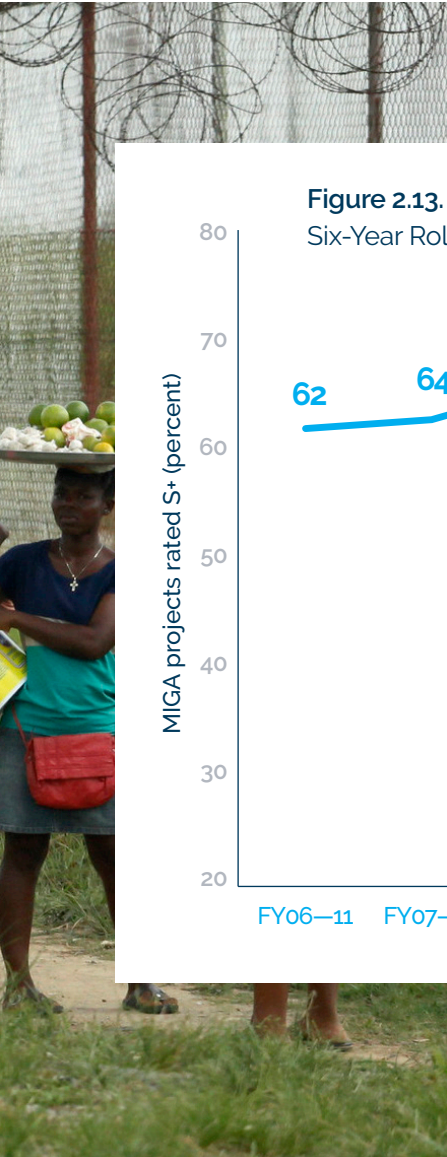
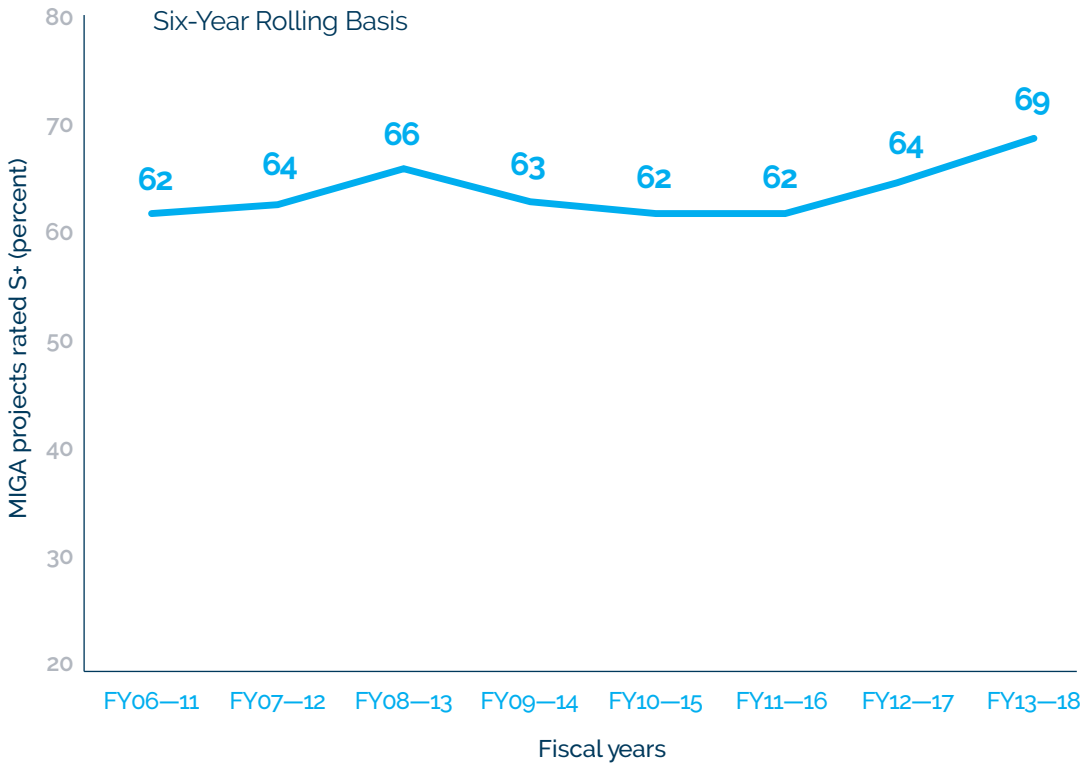


Figure 2.13. MIGA Project Development Outcome Rating
Six-Year Rolling Basis



Source: Independent Evaluation Group.

Note | FY = fiscal year
MIGA = Multilateral Investment Guarantee Agency
S+ = satisfactory or better

The financial markets sector had the lowest ratings at 58 percent S+. These low ratings for financial markets projects were caused by adverse impacts that the global financial crisis had on financial markets in Eastern Europe and Central Asia and by issues with MIGA's assessments, underwritings, and monitoring. MIGA has diversified its portfolio away from financial markets to other sectors in Eastern Europe and Central Asia, which has helped improve MIGA's performance trend in that region.

MIGA's work quality has improved. Ratings for MIGA's Assessment, Underwriting and Monitoring increased from 54 percent S+ in FY07–12 to 59 percent in FY13–18. Ratings for the environmental and social effects of MIGA guarantee projects increased from 50 percent S+ in 2007–12 to 83 percent in FY13–18 on the heels of MIGA adopting its Performance Standards on Social and Environmental Sustainability in 2007.

MIGA's clients are larger multilateral investors, reflecting its mandate to promote cross-border investment in developing countries by providing guarantees to investors and lenders. MIGA guarantees against political risks. Other investors carry the credit risk. The relatively large size of MIGA-supported projects—on average \$109 million in gross issuance—makes these projects visible to host countries and motivates governments to help these projects succeed, for example, by undertaking planned regulatory reforms. MIGA originates about 62 percent of its projects from Part 1 countries.



MIGA played an active and important role in promoting private sector investment through projects in IDA and FCV-affected countries. This is based on IEG’s review of 13 MIGA projects in IDA and FCV-affected countries evaluated in FY17 and FY18. The reviewed projects were all relevant because they fit with MIGA’s and host countries’ strategic priorities. Capable international investors who introduced competitive power generation or other technologies sponsored successful infrastructure projects. Some large-scale power projects were the first of their kind in the country. MIGA helped deter political risks and resolve emerging issues, for example, on arrears payments by governments. In successful agribusiness projects, MIGA provided reinsurance for foreign direct investment in IDA countries, created new supply chains, provided trademark license agreement guarantees, and integrated farmers and others into new processing facilities, irrigation networks, or distribution networks. Generally, the agribusiness projects were socially, economically, and environmentally sustainable, and the demonstration effect encouraged future private sector participation in the sector. This highlights a main difference between successful and less successful MIGA projects: the project’s market and business sustainability.¹⁸ For example, unsuccessful power sector projects had low market and business sustainability because of lower consumer demand for power and intense competition from rival sources of power generation. In the telecom sector, some projects were unsustainable because episodes of violence or increased competition led to fewer subscribers than expected.



¹⁸ Of the 13 projects in International Development Association and fragility, conflict, and violence-affected countries evaluated in FY17 and FY18, 10 projects were rated satisfactory or better and 3 projects were rated less than satisfactory. All of the projects fit with the Multilateral Investment Guarantee Agency’s and host countries’ strategic priorities.

3. Part II

Assessing Outcome Levels

Introduction

Project and program ratings give a helpful picture of Bank Group achievement against stated objectives, but objectives set outcomes at different levels, and the line of sight to higher-order development goals varies considerably. Beneath every rating is a wealth of information about where the Bank Group is focusing its efforts and what these mean in relation to its outcome orientation. This part of the *RAP* classifies objectives according to their outcome levels and examines links between performance and outcome levels.

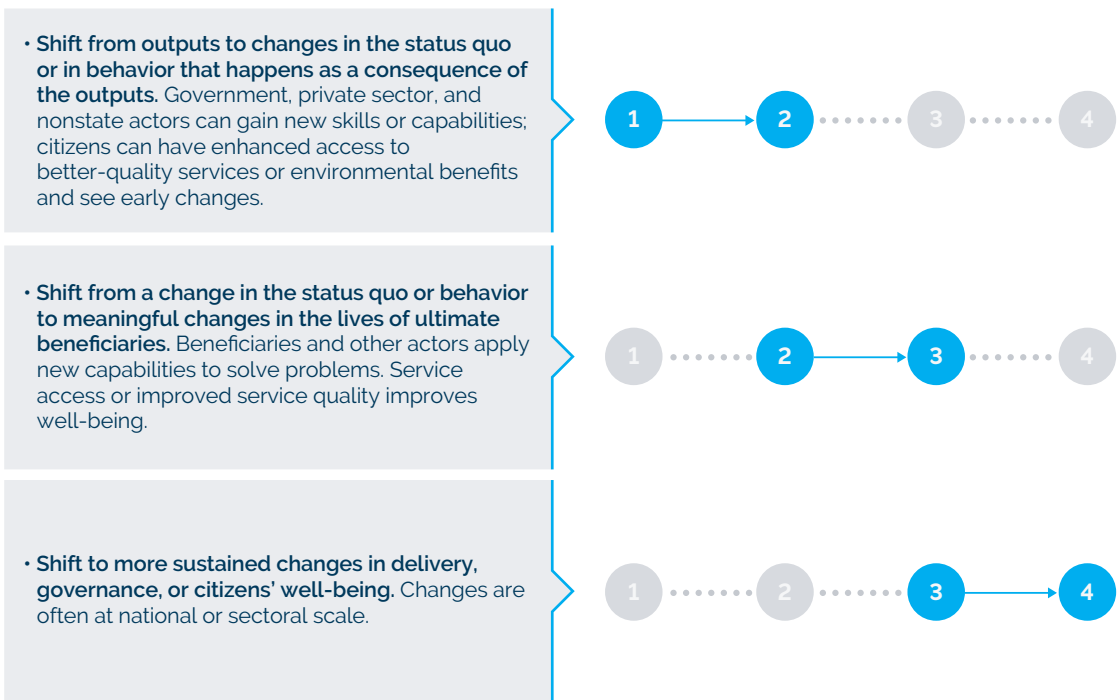


This chapter presents a theory of change framework to classify outcome levels. Because of a lack of data, the framework is by no means exhaustive, but it still offers a common lens to understand outcomes and outcome levels across sectors and Bank Group institutions, thus providing essential new information about the most typical types of outcomes that make up the project portfolio. The next section looks at the distribution of project outcome levels in samples of World Bank and IFC projects. This is followed by an assessment of the relationship between projects' outcome levels and ratings, which may help shed light on some of the risk-return trade-offs when project teams are formulating project objectives. This part concludes by reviewing the outcome orientation of key thematic areas.

Outcome Classification Framework

The novel outcome classification framework uses a theory of change logic to define comparable and complementary outcome levels. IEG synthesized sectoral theories of change derived from World Bank and IFC projects, among other sources, to build the outcome classification framework and validated the classifications on World Bank and IFC projects. Box 3.1 describes the framework and the samples that IEG applied it to. The framework defines four outcome levels. Each level corresponds to a step in a theory of change for how the Bank Group’s work influences clients’ development outcomes, ranging from outputs at level 1 to early, intermediate, and long-term outcomes at levels 2 to 4. IEG defined *shifters* to distinguish one outcome level clearly from another (figure 3.1).¹⁹

Figure 3.1. Steps in the Outcome Levels



¹⁹ These terms come from standard evaluation and results-based management literature. However, although these terms call attention to outcomes' time dimension, the coding framework emphasized the sequential steps in the logic of how interventions lead to outcomes.

For example, level 1 outcomes include project deliverables, but level 2 outcomes include changes in the development status quo that resulted from the level 1 deliverables. Hence, level 2 outcomes follow quite directly from project outputs and often focus on improved access, capacity, regulation, planning, provision, and quality of public services—all of which represent relatively immediate benefits to beneficiaries. Level 3 outcomes follow indirectly from project interventions and are beyond the direct control of the World Bank and its clients. At level 3, the level 2 outcomes have led to material improvements that solved development problems, causing sectorwide ripple effects that benefit end beneficiaries. The ripple effects of level 4 outcomes are even deeper and wider. These are outcomes with systemic effects nationally or across sectors that contribute to general well-being. Level 4 outcomes correspond to the Sustainable Development Goals, the twin goals, and other higher-level outcomes to which the Bank Group aspires. Figure 3.2 shows representative examples taken from World Bank project objectives.²⁰

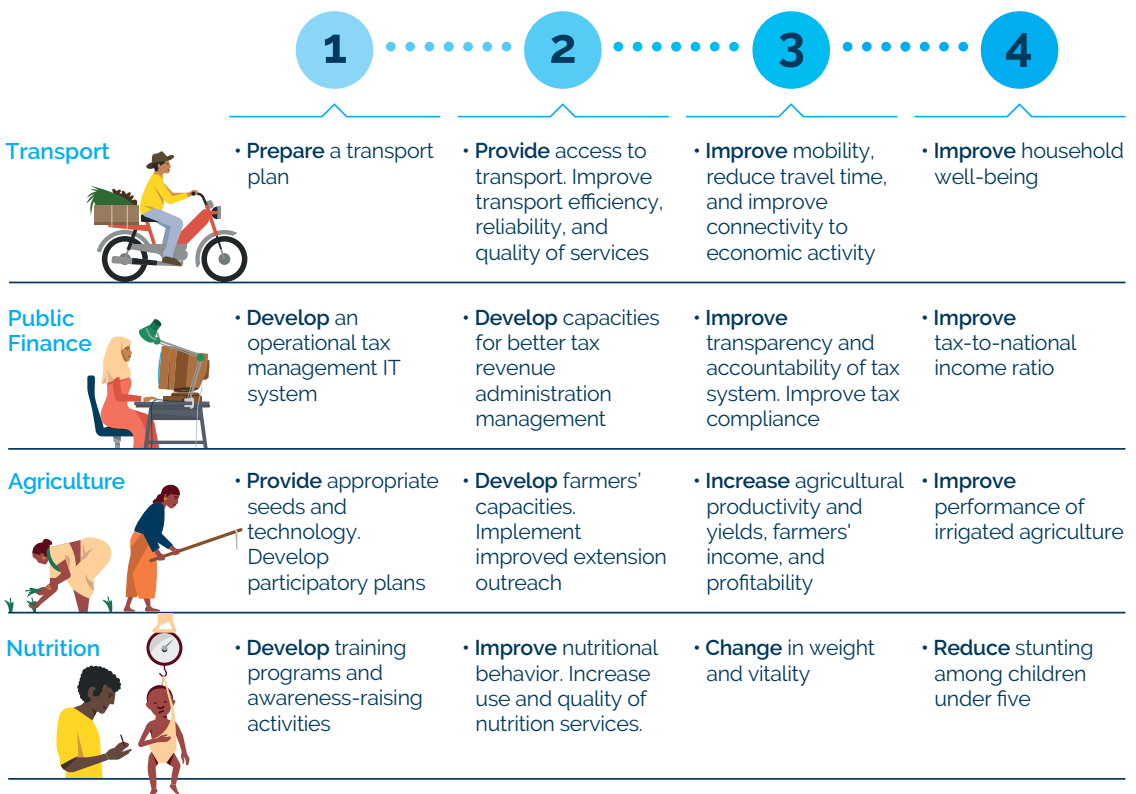
The framework captures World Bank projects' intended and achieved outcomes and IFC projects' intended outcomes. IEG designed the framework to compare outcomes in a comparable



²⁰ This is a departure from the IEG's traditional project ratings. These are determined by IEG after projects close, based on achieved outcomes relative to intended objectives. IEG project development objective ratings consider only the declared project development objective to a limited extent, namely, in the assessment of the project's relevance.

manner across sectors (box 3.1). At project design, all project documents state a clear objective, called a project development objective at the World Bank and called claims at IFC under its new Anticipated Impact Monitoring and Measurement framework. During project implementation, teams and clients manage projects to achieve these objectives. When projects close, self-evaluations review whether they achieved their stated objectives, with validation by IEG.

Figure 3.2 Examples of Outcome Levels



Source: Independent Evaluation Group

Note | GDP = gross domestic product | IT = information technology

Box 3.1. The Outcome Level - Framework

This framework and its application have strengths and weaknesses. Comparability across sectors and countries is a key strength, which the Independent Evaluation Group (IEG) ensured by defining comparable yardsticks and applying internal and external quality assurance. For example, IEG validated the framework through pilots and expert consultations, defined key terms in project development objectives (PDOs) that are indicative of different outcome levels, developed detailed coding guidance, and tested for interrater reliability (the reliability of multiple coders to code the same outcomes) by having multiple team members independently code the same PDO to standardize coding scores. However, the framework is a blunt tool. It focuses on stated and measured objectives, which may not be the same as the actual outcomes. It simplifies outcomes' complex social realities into four categories that do not factor in context, so one country's simple achievement could be another's ambitious outcome.



Coding focused on PDOs, which are summaries that approximate projects' intended outcomes, but sometimes they are vague or may not comprehensively reflect all of the project's objectives. To overcome this challenge, IEG consulted the project's indicators when in doubt but did this less often for investment project financing than for development policy financing, which was harder to assess because of long PDOs with multiple parts. For composite PDOs with more than one subobjective, IEG chose the highest. For two samples, new World Bank projects and new International Finance Corporation (IFC) projects, IEG followed a different approach that reviewed PDOs and indicators' outcome levels separately to compare them. Another challenge of the framework was differentiating between level 3 and level 4 outcomes. The difference between these outcomes is conceptually clear, but in practice, it can require the coder to subjectively judge the projects' real objectives and then approximate how deep and how systemic are the outcomes to which these objectives aspire.

IEG applied the framework to four project samples:

- Recently closed projects: all 989 Implementation Completion and Results Report Reviews completed from fiscal year (FY)17 to FY20 (April). This sample is large enough to allow Global Practice comparisons.
- Older projects with IEG field evaluations: all 42 Project Performance Assessment Reports from FY19 and FY20 available in March 2020. Analysis focused on achieved outcomes in the sample's 114 component objectives. This sample shows actual project outcomes that IEG verified in the field.
- New World Bank projects: a statistically representative sample of 161 projects approved in FY19, indicative of recent approvals.
- New IFC projects: a random sample of 29 recently approved IFC investment projects. Analysis covered the 100 project and market claims in this sample.^a



Source: Independent Evaluation Group.

Note | ^a The Independent Evaluation Group assessed project objective statements and related indicators in International Finance Corporation projects approved in FY20 to understand the types and levels of outcomes in projects processed under its new Anticipated Impact Monitoring and Measurement (AIMM) system. IEG identified all AIMM claims in the project summaries for 29 randomly selected investment projects and indicators in 21 of these projects. This was not an evaluation of AIMM as a tool. IEG did not review AIMM scores or the underlying methodologies for calculating AIMM scores or review projects' actual outcomes. The 29 sampled projects made 100 claims, of which 61 were project claims and 39 were market claims. All project and market claims were clearly formulated objective statements. Sampled projects contained 142 indicators, of which 58 percent were project indicators, 18 percent were market indicators, and the remaining 24 percent required corporate and other indicators unrelated to AIMM claims.

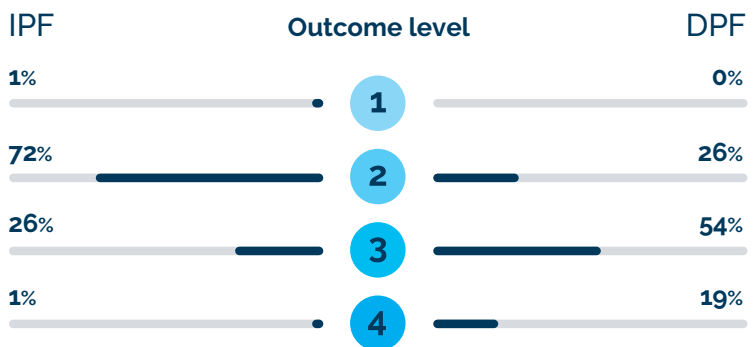
Project Outcomes

This section analyzes the distribution of projects' objectives to understand what types of outcomes most projects intend to achieve and measure. It does so partly in response to Board members' demands for more evidence on outcomes. Until now, the understanding has been that projects pursue diverse objectives across diverse sectors, contexts, and instruments, with limited room for generalization. For the first time, this research shows that project objectives cluster in clear patterns depending on sector and lending instrument. Most IPF objectives cluster at level 2 around quality and access to services. A few sectors, most notably agriculture and environment, state IPF objectives at level 3 with clearer focus on end beneficiaries, and most DPFs state their objectives at level 3 with a focus on policy reform outcomes. Recently approved IFC projects often state their objectives at level 3, particularly in relation to market creation objectives.

Most IPFs have project development objectives that aim for level 2 outcomes. IEG classified 72 percent of IPF projects' objectives in the recent Implementation Completion and Results Report Review (ICRR) sample at level 2 (figure 3.3).²¹ By far, the most common level 2 IPF objectives improve quality or access to social- or infrastructure-related public services. Most IPFs—and by extension most of the World Bank's work—intend to strengthen public sector capacity. This reflects the strong emphasis in World Bank operations on improving public sector capacity and performance as an enabler of higher-level change. The prevalence of service access objectives also reflects the relatively easier measurement and attribution to World Bank support of such objectives.



Figure 3.3. Outcome Levels in IPF and DPF Projects



Source: Independent Evaluation Group.

Note | DPF = development project financing | IPF = investment project financing

²¹ The share was similar in the recently approved sample, at 68 percent of investment project financing objectives at level 2.

IPFs in a few sectors pursue level 3 outcomes more often. Level 3 outcomes were found in 26 percent of all IPF project objectives and level 4 in 1 percent (figure 3.3). However, there is clustering in some sectors: half of Agriculture and Environment GP projects, 35 percent of Transport GP projects, 31 percent in the Water GP, and 27 percent in Energy and Extractives. The share of level 3 and 4 objectives is far lower in other GPs, ranging between 10 and 14 percent. Common examples of IPF level 3 outcomes include improved agricultural productivity, yields, and incomes; improved management of protected areas; climate resilience; and transport connectivity. A focus on sectorwide change and end beneficiaries characterizes these types of level 3 outcomes. Such outcomes are different from most IPFs' focus on level 2 service access and capacity. The reason for the variation across GPs in objectives' outcome levels is not entirely clear, though the ability to define suitable indicators plays a role in how teams set objectives.

DPFs' objectives cluster around yet other types of outcomes. Objective statements at outcome levels 3 and 4 were found in 54 and 19 percent, respectively, of World Bank DPFs (figure 3.3). DPFs seek to induce change through policy, institutional, and governance reforms. DPFs achieve their objectives less often, resulting in lower ratings compared with IPFs, as seen in part I. Representative examples of level 3 DPF objectives include macrofiscal stability, improved transparency and accountability, and increased domestic tax revenue. However, 26 percent of DPF outcomes in the ICRR sample were at level 2. Examples include technical support to policy and regulatory reforms and the first of a series of planned DPFs, with higher intended outcomes for subsequent DPFs.

Level 1 outputs are rare in project objective statements. Only 1 percent of objectives in recent ICRRs and 9 percent of objectives in recent approvals had level 1 outputs in the objective statement. No DPFs or IFC projects in the samples had output objective statements. It is established good practice to focus on outcomes, so it is positive that so few projects have level 1 objective statements.

Project objectives in FCV-affected countries are not distributed differently. In FCV projects, 71 percent of objectives are at level 2, 25 percent are at level 3, and 4 percent are at level 4. This compares with 64 percent at level 2, 31 percent at level 3, and 4 percent at level 4 in non-FCV-affected countries. The similarity of outcome levels in FCV and non-FCV countries is surprising because of the higher contextual risks in FCV-affected countries and the need for quick and simple attainable goals, as discussed in part I. IEG also observed that results



frameworks in FCV projects did not commonly capture conflict drivers or outcomes on fostering the country’s resilience to conflict and violence. FCV countries need agile responses to their unique challenges, an aspect that falls outside the outcome level classification framework.

IEG also compared projects’ indicators to their objective statements to assess whether the indicators’ levels matched the objectives’ outcome level. It found that 14 percent of recently approved projects have no indicator of the same level as the objective, which could suggest that there are no indicators able to measure the objective’s achievement.

Recently approved IFC projects under the Anticipated Impact Monitoring and Measurement system often state their objectives in relation to higher-level outcomes because they are aligned with IFC’s goals of creating markets and fostering private sector development (box 3.2). Outcome level 3 and 4 objective statements were found in 67 and 15 percent, respectively, of recent IFC market claims, and in 39 and 13 percent, respectively, of project claims (figure 3.4). Level 2 objective statements were found in 18 percent of market claims and 48 percent of project claims. Figure 3.5 shows representative examples of IFC claims. Twelve percent of IFC project and market claims did not have indicators that matched the claim’s outcome level, which could suggest that none of the selected proxy indicators are able to measure the objective’s achievement.



Figure 3.4. IFC Project and Market Claims’ Outcome Levels



Source: Independent Evaluation Group.

Note | IFC = International Finance Corporation.

Box 3.2. IFC's AIMM System for Setting Project Objectives

The Independent Evaluation Group (IEG) analyzed objectives in recent International Finance Corporation (IFC) projects to understand how IFC's new Anticipated Impact Monitoring and Measurement (AIMM) system articulates intended outcomes.^a Under AIMM, IFC projects include multiple project claims and market claims, but the World Bank sets only one objective per project. Project claims are defined as a project's direct and indirect effects on stakeholders, the economy, and the environment and are comparable to World Bank projects' project development objectives. Market claims are derived effects, defined as a project's ability to catalyze systemic changes beyond those effects brought about by the project itself. IEG did not review AIMM scores, an index number for a combination of the depth and likelihood of project outcomes and contribution to market creation.

Overall, IEG found that the system of project and market claims contained clear objective statements that aligned well with IFC's higher-level goals of creating markets and fostering private sector development. AIMM ensures that project objectives align with IFC's goals. Although IFC can ensure such alignment because of its focused business model and goals, the World Bank operates with objectives that are more diverse because of its diverse sector and country contexts. It is too early to tell what impact AIMM will have on outcome achievement, ratings, evidence, and incentives because no project under AIMM has been evaluated yet.

Source: Independent Evaluation Group.

Note | ^a This sample is different from the rated sample analyzed in part I, which did not include projects with AIMM claims.

Figure 3.5. Representative Examples of IFC Claims



Source: Independent Evaluation Group

Note | IFC = International Finance Corporation.

Project Outcome Levels and Ratings

This section combines the outcome level classification and ratings to examine the relationship between projects' outcome levels and projects' performance. This analysis was motivated by World Bank management's efforts to better identify the risk-return trade-offs when formulating project objectives and related questions about whether the rating system influences project teams' incentives when setting objectives. The analysis also aimed to explore potential explanations for key performance patterns identified in part I. Efficacy ratings (which assess to what extent projects achieve their stated objectives) and outcome ratings (which consider the project's relevance and efficiency) were used.²²

The relationship between objectives' outcome levels and projects' performance is only modest and becomes insignificant when controlling for other factors. Specifically, ratings for projects with level 3 and 4 outcomes are modestly lower than for projects with level 2 outcomes, and the difference in ratings is insignificant when controlling for instrument and other factors such as M&E quality (box 3.3). This finding runs counter to a key assumption prior to doing the analysis that one of the reasons for not setting higher level objectives is the risk of a lower rating. Instead, the finding shows no systematic trade-off between projects' outcome level and ratings. This implies that many projects with higher-level objectives manage to achieve good outcome ratings, in part by having strong results frameworks to measure outcome achievement. Although the model does not provide any more detail on the causal relationship between objectives set at design and projects' eventual performance—both depend on specific country and sector contexts—it does point to larger questions about when it makes sense for projects to set higher-level objectives and what it takes for such projects to be successful in reaching their intended outcomes.

²² IEG uses a numerical conversion of the four-point efficacy rating. Efficacy ratings are sometimes given for subobjectives. In that case, the average of the subobjective ratings was calculated.



Box 3.3. Regression of Projects' Performance on Outcome Levels and Other Factors

A regression analysis on the Implementation Completion and Results Report Review sample shows that projects' outcome levels do not play a statistically significant role for these projects' efficacy rating when controlling for lending instrument. Regressing efficacy ratings on outcome levels and a dummy for lending instrument shows that investment project financing projects have markedly higher efficacy ratings than development policy financing projects do, in line with the findings of part I (model 2 in table B3.3.1). The difference in efficacy rating between lending instruments is statistically significant at the 0.001 percent level, whereas the outcome level is not statistically significant in this model. The negative relationship between efficacy and outcome level in model 1 is driven by the fact that investment project financing, which have higher efficacy ratings than development policy financing, also have lower outcome levels. The results are also robust to including projects' monitoring and evaluation quality rating (model 3). In this model, the lending instrument and monitoring and evaluation quality affect efficacy rating at the 0.001 significance level, and outcome level remains statistically insignificant. The results are the same when controlling for Global Practice as random effect (model 4).

Table B3.3.1. Regression Results

Variable	1	2	3	4
Outcome level	-0.0970** (0.0323)	-0.0500 (0.0356)	-0.0305 (0.0298)	-0.0305 (0.0328)
IPF (vs. DPF)		-0.1982*** (0.0538)	-0.2151*** (0.0427)	-0.2151*** (0.0304)
M&E rating			-0.4793*** (0.0243)	-0.4793*** (0.0328)
GP as random effect				Yes
Number of observations	949	946	944	944
R2	0.0102	0.0225	0.3201	0.3201

Source: Independent Evaluation Group.

Note | DPF = development policy financing
| GP = Global Practice

| IPF = investment project financing
| M&E = monitoring and evaluation

**p > .01
***p > .001

Pairwise comparisons of efficacy and outcome ratings illustrate the same tendency of modestly lower ratings as outcome levels increase (figure 3.6 and table 3.1). DPFs at level 3 are rated modestly lower on outcomes than DPFs at level 2—true in both the ICRR and the Project Performance Assessment Report sample—and DPFs at level 4 are rated lower. Efficacy ratings are marginally lower for DPFs at higher outcome levels. IPFs tell a similar story. IPFs with level 2 outcomes have marginally higher efficacy ratings and somewhat higher outcome ratings than IPFs with level 3 outcomes—77 percent MS+ compared with 72 percent MS+. (The result for level 4 is not robust because of the small sample size.) Similar patterns are seen in many GPs and in a large sample of older projects (box 3.4).²³ IEG next examines whether outcome levels help explain ratings differences reported in part I between GPs and project types.

Table 3.1. Ratings and Outcome Levels, by Instrument

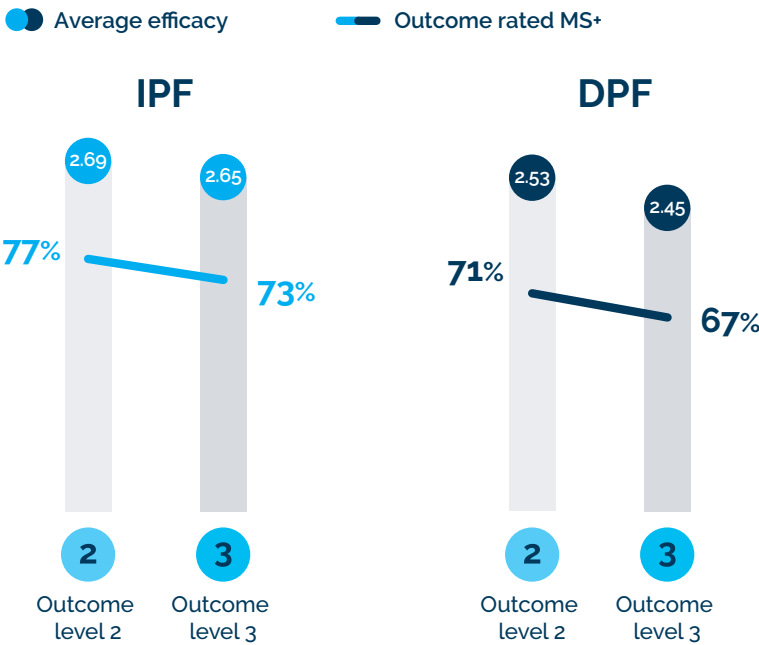
Outcome Level	IPF			DPF		
	Projects (no.)	MS+ (percent)	Average efficacy rating ^a	Projects (percent)	MS+ (percent)	Average efficacy rating ^a
Level 1	9	89	2.8	0	n.a.	n.a.
Level 2	580	77	2.7	45	71	2.5
Level 3	211	73	2.7	93	68	2.5
Level 4	7	57	2.6	33	61	2.3
Total	807	76	2.7	171	67	2.5

Source: Independent Evaluation Group.

Note | DPF = development policy financing | MS+ = moderately satisfactory or above
 | IPF = investment project financing | n.a. = not applicable
^a IEG uses a numerical conversion of the four-point efficacy rating.

²³ For example, Transport projects with level 3 outcomes have somewhat lower outcome ratings (74 percent moderately satisfactory or above) than projects with level 2 outcomes (83 percent moderately satisfactory or above), though efficacy ratings are identical for both outcome levels: 2.7 out of a maximum of 4.

Figure 3.6. Ratings and Outcome Levels, by Instrument




Source: Independent Evaluation Group.

Note | Circles show efficacy ratings, and lines show percentage of projects rated MS+.
 DPF = development project financing | IPF = investment project financing
 MS+ = moderately satisfactory or above.

Box 3.4. Outcome Levels and Ratings over a Longer Period

The Independent Evaluation Group used machine learning to extend the outcome classification to older projects. It used the population of all 3,119 projects that were rated since 2009 for which the relevant information was readily available. The machine learning algorithm classified projects based on their objectives at either level 2 or level 3, with 92 percent accuracy on a test data set. Precision was lower for levels 1 and 4 because of small sample sizes, and therefore these results are not used here. Looking at only investment project financing (IPF), the outcome levels were broadly constant across the years. The algorithm coded 75 percent of projects at level 2 and 25 percent at level 3. Development policy financing projects had higher outcome levels than IPF projects in the machine-coded data. Furthermore, consistent with the other findings of this section, the outcome ratings for level 3 IPF projects were only marginally lower than level 2 IPF projects—71 percent moderately satisfactory or above compared with 73 percent.

Source: Independent Evaluation Group.



When revisiting the key performance patterns identified in part I, IEG finds that projects' outcome levels do not explain the low ratings for projects in the MTI and Governance GPs. IEG combined the Education; Urban, Resilience, and Land; and Transport GPs (which tend to deliver basic services and are among the highest-rated GP portfolios) and combined the MTI and Governance GPs (both of which focus on policy and institutional reforms, often using DPFs, and are the lowest-rated GP portfolios). Table 3.2 shows that MTI and Governance projects have lower ratings compared with Urban, Education, and Transport projects, regardless of their outcomes level. The table also shows that MTI and Governance projects with level 3 outcomes achieved the same ratings as projects with level 2 outcomes (61 percent MS+ compared with 60 percent), and there was only a limited ratings decline for projects with level 4 outcomes (55 percent MS+). Looking only in FCV-affected countries, MTI and Governance projects with level 2 and 3 outcomes are again rated equally. Instead, the explanation for these key performance trends is related to the DPF instrument, which MTI and Governance GPs use much more often than other GPs.

Multiple factors help explain lower ratings for DPFs (and thus for MTI and Governance GP projects). Policy and institutional reform objectives are more prone to risk and uncertainty than service delivery objectives. Some of those risks relate to the longer time frame needed for DPFs' policy reforms to lead to outcomes. Such reforms must successfully proceed through a long change pathway to arrive at desired outcomes. For example, a policy reform supported by a DPF must build from a prior action (for example, a parliamentary proposal for a legislative change) to approving and enacting the change and waiting for that change to achieve intended higher-level outcomes, such as people or firms behaving differently and spurring economic growth. Each of the links in this chain depends on actions by governments, parliaments, and economic actors outside of the project's control. Some risks relate to the nature of the DPF instrument itself. For example, the World Bank has less room to make course corrections to achieve results in DPFs than it does in IPFs, especially in stand-alone DPFs. Evaluation methods also play a role because, in reality, they differ between IPFs and DPFs. For example, DPFs' outcome ratings are based only on assessment of relevance and efficacy, with no assessment of efficiency as done for IPF, and furthermore with challenges in assessing DPFs' relevance and efficacy.²⁴

²⁴ It is hard to assess how and how much development policy financing contributed to overall reform outcomes, given that development policy financing's prior actions are part of broader reform plans. Instead, evaluators can focus on the relevance of the prior actions and the results indicators. Planned reforms of the evaluation methodology for development policy financing aim to strengthen these dimensions.

Table 3.2. Ratings and Outcome Levels for Select Global Practices and Project Types

Outcome Level	Education, URL, and Transport			MTI and Governance		
	Projects (no.)	MS+ (percent)	Average efficacy rating ^a	Projects (percent)	MS+ (percent)	Average efficacy rating ^a
Level 1	6	100	3.1	0	n.a.	n.a.
Level 2	237	84	2.8	53	60	2.5
Level 3	39	77	2.7	100	61	2.4
Level 4	3	100	2.8	28	55	2.2
Total	285	84	2.8	181	60	2.4

Source: Independent Evaluation Group; recent Implementation Completion and Results Report Review sample.

Note | MS+ = moderately satisfactory or above | URL = Urban, Resilience, and Land
 | MTI = Macroeconomics, Trade, and Investment | n.a. = not applicable
 | ^a IEG uses a numerical conversion of the four-point efficacy rating.

Projects with level 3 and 4 objectives appear to have adequate result frameworks and M&E systems as often as other projects do. It seems intuitive that it would be harder to design adequate result frameworks for projects with higher-level outcomes, yet the evidence suggests otherwise.²⁵ Project M&E ratings decline little as outcome levels increase. Projects with objectives at level 2 were rated 45 percent high or substantial on M&E quality compared with 43 percent for level 3 and 44 percent for level 4 projects (table 3.3). A similar pattern emerges when looking at IPFs only. Similarly, IEG rates IPF projects low when there is insufficient evidence to confirm the projects’ achievement of objectives. This happened to at least 6 percent of all IPF projects with level 2 outcomes and 9 percent with level 3 outcomes (table 3.3).²⁶

²⁵ Recall that the quality of projects’ monitoring and evaluation is important for ratings, according to the regression analysis and the analysis presented in part I.

²⁶ These figures are a lower bound estimate based on Implementation Completion and Results Report Reviews, in which the IEG reviewer explicitly noted weak evidence as a reason for the rating decision.

Table 3.3. Ratings and Outcome Levels for Select Global Practices and Project Types

Outcome Level	Projects Rated High and Substantial on M&E	IPF Projects Rated High and Substantial on M&E	IPF Projects with Lack of Evidence
Level 1	n.a.	n.a.	n.a.
Level 2	45	45	6
Level 3	43	41	9
Level 4	44	n.a.	n.a.

Source: Independent Evaluation Group.

Note | Values for sample sizes of 10 or fewer projects not shown. | IPF = investment project financing | n.a. = not applicable
 | M&E = monitoring and evaluation

The risk-return trade-off does not appear to be very pronounced in these data. Outcome levels vary across GPs and instruments, but this is not the reason for performance differences because IPF projects with level 3 objectives and DPF projects with level 3 and 4 objectives do not appear to have markedly higher risk of weak performance compared with projects with lower-level objectives. Half of agricultural and environmental IPF projects set their objectives at level 3 and still register mostly strong achievements. Differences in performance appear to be more closely associated with levels of risk and uncertainty and the time and complexity involved in pursuing policy and institutional reforms. Questions remain about what is required for projects to set and achieve ambitious objectives.



Thematic Area Outcomes

This section considers how the Bank Group aggregates project and program results in key thematic areas and the implications for its outcome orientation. The IEG team reviewed corporate strategies, Corporate Scorecards, and results measurement systems for three Global Themes: Gender; Climate Change; and Fragility, Conflict, and Violence (looking particularly closely at Climate Change).

The Bank Group has clearly articulated higher-level outcomes for its thematic work. Bank Group corporate strategy documents set out clear high-level outcome goals, most famously the twin goals on poverty and shared prosperity and the commitment to the Sustainable Development Goals. Furthermore, there are many other goals, targets, and policy commitments set in different sectoral and thematic areas through the 2018 Bank Group capital package; the IDA Replenishments; *World Bank Group Climate Change Action Plan 2016–2020*; *World Bank Group Gender Strategy (2016–2023): Gender Equality, Poverty Reduction, and Inclusive Growth*; and *World Bank Group Strategy for Fragility, Conflict, and Violence 2020–2025*, among others.

The Bank Group has extensive systems to track and aggregate its results, but these systems often operate at some distance from higher-level outcomes. All projects and country programs have results frameworks with objectives, indicators, and M&E systems to capture those indicators. These projects and programs undergo self-evaluations that IEG validates and rates, and these form the backbone of the Bank Group’s results measurement system. Aggregated data from projects and country programs appear in the Bank Group’s Corporate Scorecards, IDA’s results measurement system, and thematic results measurement systems, such as those for gender and climate change. Yet these data focus on internal processes and the number of people reached by health, water, financial, education, sanitation, electricity, and agricultural services. Such reach indicators correspond to level 2 outcomes, but they convey little about the service’s quality and impact on human well-being and, therefore, do not help staff manage to those outcomes. Only a few of the indicators in the Corporate Scorecards, IDA’s results measurement system, and thematic results measurement systems track higher-level outcomes.

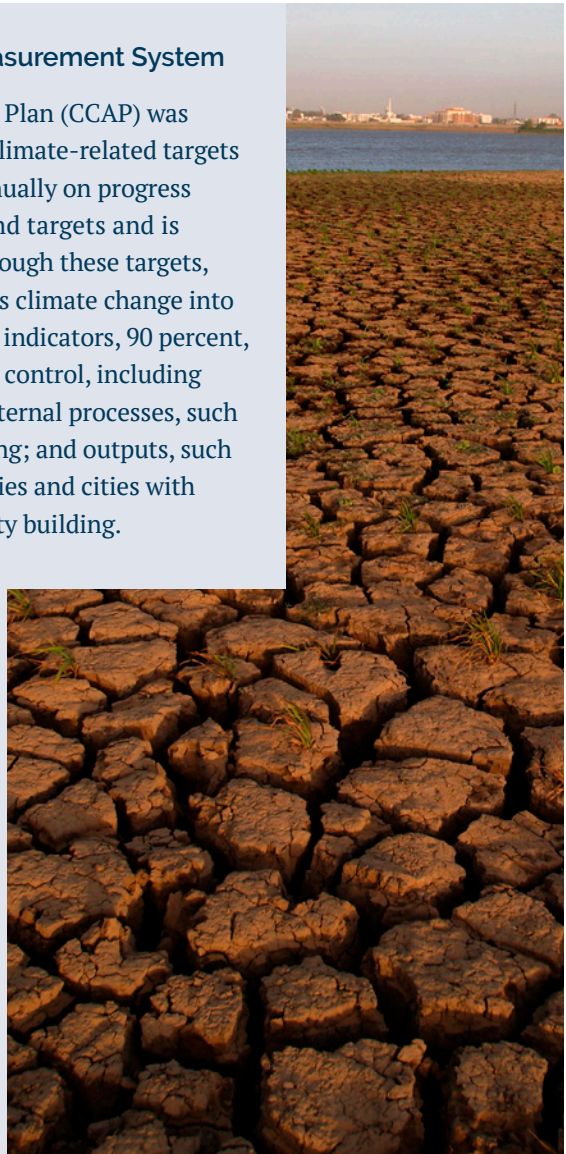
The results measurement systems for thematic areas do little to support the Bank Group’s outcome orientation. The RAP defines *outcome orientation* as gathering credible evidence on outcome achievement; using this evidence to adapt interventions and portfolios, engage clients, and learn; and thus becoming more effective at achieving positive social change. This definition is not about encouraging staff to aim for any particular level of outcomes. Rather, strong outcome orientation requires collecting credible evidence on progress and achievements and ensuring that staff have the right incentives to use the evidence to pursue positive social change relevant to the context of countries and sectors. Outcome orientation is different from achieving targets and monitoring processes.

Instead, corporate results measurement systems help senior management track and incentivize operational fulfillment of corporate policy commitments. The Bank Group’s ability to track and report on its policy commitments confers legitimacy and credibility on the organization and has undoubtedly helped it secure strong IDA replenishments and IBRD and IFC capital increases. Corporate indicators incentivize operations to integrate these themes into their work streams and meet targets. For example, when the World Bank committed to engage citizens in all applicable projects and started tracking this, the share of projects with citizen engagement indicators in their results frameworks increased quickly, but there was limited evidence on the quality, influence, or outcomes of citizen engagement (World Bank 2018a). Box 3.5 examines how the climate change results measurement system has helped the Bank Group meet or exceed its climate action targets.

Box 3.5. The Climate Change Results Measurement System

The World Bank Group Climate Change Action Plan (CCAP) was adopted in April 2016 and lays out ambitious climate-related targets for 2016–20. The Bank Group has reported annually on progress for over 30 climate change–related actions and targets and is preparing a retrospective summary report. Through these targets, the Bank Group monitors how well it integrates climate change into operations and strategies. The vast majority of indicators, 90 percent, relate to actions under the World Bank’s direct control, including inputs, such as financing for climate action; internal processes, such as greenhouse gas accounting and risk screening; and outputs, such as the number of products that support countries and cities with climate-related policies, strategies, and capacity building.

The results measurement system used for tracking CCAP targets has a limited focus on projects’ and programs’ quality and higher-level outcomes. The system incentivizes operations to adhere to process requirements, for example, to adjust cost-benefit analysis for the shadow price of carbon and screen or assess projects’ potential climate risks. However, it is unclear if requiring risk screening influences projects’ and country programs’ design, quality, and outcomes. Furthermore, there is no evidence on how well projects address climate risks.



In the CCAP itself, the main commitment to increase the share of climate change–related commitments to 28 percent has driven all subsequent outputs and outcomes. At the level of the many institutional CCAP targets, only approximately 10 percent relate to outcomes, including level 2 outcome indicators, such as the amount of commercial funds mobilized for clean energy or the number of people covered by climate-adaptive social protection and early-warning services.

The CCAP reporting does not, however, assess the Bank Group’s contributions to greener or more resilient national development trajectories. On the whole, the CCAP results measurement system has driven accountability and internal incentives to mainstream climate action across the Bank Group and has tracked progress in meeting targets, but it does not guide operations toward key outcomes or assess the quality of those outcomes.

Source: Independent Evaluation Group.



Corporate mandates and indicators cascade down to operational departments and can potentially drive box-checking behaviors. If operations sought to maximize the reach indicators for service access in the Bank Group Corporate Scorecard, they could increase the number of people covered by water, health, electricity, and other services at the cost of service quality. However, if operations instead had evidence of the quality of services, the capacity of institutions, and beneficiaries’ productivity and well-being, they might be better able to manage for those outcomes. In another example, an emergency health project in an Ebola-affected country was held back at one point because it did not meet a minimum threshold for climate cobenefits. Overall, the challenge is to ensure that targets create incentives that are compatible with outcome orientation, as discussed in the next, concluding, chapter.




4. Conclusions

Getting to Outcomes

This report expanded on past *RAPs* by focusing not only on core performance as assessed through ratings but also on outcomes and relationships between ratings and outcome levels. This chapter draws out some key findings, conclusions, and implications for the Bank Group's COVID-19 response and then for its outcome orientation. It finds that there are trade-offs between using results measurement systems for tracking commitment targets and outcome orientation. Confronting these trade-offs is necessary if the Bank Group wants to better support outcome orientation.

Findings and Conclusions

The analysis of performance showed positive ratings trends for World Bank and MIGA projects and Bank Group country programs in IBRD countries. The analysis linked the positive outcome ratings trends to strong work quality on project design, implementation support, and M&E, and broadly conducive economic and institutional conditions in many larger countries before the pandemic. Performance trends for IFC projects and in FCV-affected countries are less positive, albeit with signs of recent slight improvements for IFC. Less successful results were often linked to large shocks and issues with projects' and programs' preparation for risks and their response when shocks occurred.




The analysis of outcomes showed that most IPF objectives cluster at level 2 around quality and access to services, though a few sectors state IPF objectives at level 3 with a clearer focus on end beneficiaries. Most DPFs state their objectives at level 3 with a focus on policy reform outcomes, and recently approved IFC projects often state their objectives at level 3, particularly in relation to market creation objectives. Projects' outcome levels have only a modest relationship with their ratings, and the relationship becomes insignificant when controlling for other factors. Looking beyond projects, the analysis showed limited higher-level outcome data. The existing results measurement systems collect evidence needed for ratings and for process and compliance monitoring, which is different from evidence on outcome achievement.

Based on the evidence and findings, this *RAP* concludes that the Bank Group can improve how its incentives and results measurement systems support outcome orientation. At the project level, many projects with higher-level objectives manage to achieve good IEG ratings, in part by having strong results frameworks to measure outcome achievement. Even so, it would not be realistic or desirable to expect all World Bank projects to have objectives at outcome level 3 or 4, as discussed in box 4.1. At the country program level, the Bank Group has opportunities to take a broader and more strategic view beyond individual projects, yet there is often little evidence on higher-level country outcomes, as discussed in IEG's forthcoming evaluation of country programs' outcome orientation. At the corporate level, the Bank Group's extensive systems cover different thematic work areas and collect process and output indicators to help senior management incentivize and report on operations' fulfillment of corporate policy commitments, but they do not help staff to manage for higher-level outcomes.

Box 4.1. Setting Project Objectives

Objective setting needs to balance the opposing demands of realism and ambition. Realism demands that objectives be achievable, given the projects' resources, timeline, and context. Objectives that are far removed from project interventions jeopardize the ability to show contribution. Outcome achievement should be measurable. Country and sector context, geographic scope, and beneficiaries are some of the context factors that also matter. Ambition, however, demands objectives with a line of sight to systemwide, transformative, or other important higher-level outcomes. Ambition also demands result frameworks that measure how the project changes beneficiaries' conditions, with attention to gender and distributional aspects. Balancing realism and ambition requires judgment and dialogue between client counterparts and the World Bank. Therefore, universal rules are unlikely to be helpful. In practice, it is plausible that some projects aspire to and achieve outcomes at a higher level than those captured in their objectives.

Source: Independent Evaluation Group.



Excessive focus on monitoring targets can cause a risk-averse corporate culture and stifle staff's intrinsic motivation to pursue positive social change. IEG's evaluation of the Bank Group's self-evaluation systems found that staff often have little use for the collected data and find little value in it. Instead, incentives are to focus on checking the box, meaning meeting targets and feeding the demands for corporate monitoring data (World Bank 2016a). At the same time, a corporate culture focused on compliance, disbursements, and meeting targets can induce risk aversion, reduce openness about problems, interfere with staff learning and experimentation, and stifle how staff use evidence to pursue outcomes.²⁷ For these and other reasons, there are trade-offs between outcome orientation and using results measurement systems for reporting and incentivizing fulfillment of policy commitments.

Confronting trade-offs related to the purposes of the Bank Group's results measurement systems is necessary for improving outcome orientation. The corporate results measurement systems for projects, programs, and thematic areas were purposefully designed to meet the Bank Group's need to collect data it can report to shareholders to show attributable results and that allow shareholders' representatives to hold it accountable. The purpose of tracking and reporting results data dictates what systems collect, leading to a focus on tracking commitments using quantifiable indicators of activities and lower-level results that can be attributed to Bank Group interventions, added up across portfolios, and used to verify whether targets have been met. Systems designed for such purposes are not geared to ward understanding and managing higher-level outcomes.

Implications

This *RAP's* findings and conclusions have implications for the Bank Group's ongoing response to the pandemic and other shocks. Some projects may need more robust implementation support and more frequent course correction during implementation, both to respond to shocks and unforeseen circumstances and to counter the potential effects of short preparation times on quality at entry. M&E systems across the Bank Group need to enable such responses and can do so by maintaining sight of project objectives and enabling teams and clients' identification of issues and nimbler course corrections. The World Bank's and clients' administrative procedures for restructuring and canceling projects could be streamlined. Rating systems should not unduly penalize necessary changes to targets and objectives set at approval that may need adjustment in light of COVID-19 and other shocks. Box 4.2 explains what that might involve for IFC. Furthermore, the analysis of IFC's performance highlighted the need for enhanced tools and processes to identify and mitigate market, country, and sponsor risks.

²⁷ Based on research covering many development agencies (including the World Bank), academic Dan Honig discusses how to promote staff's intrinsic motivation to achieve outcomes. Arguing for more "navigation by judgment," Honig suggests promoting a less risk-averse corporate culture that embraces bold ambitions and gathers, uses, and learns from outcome evidence (Honig 2018, 2020).

The analysis of country programs in low-capacity countries suggests that when adding new elements in response to large shocks, there is a need to simplify other program elements to avoid overtaxing country capacity. More broadly, such programs need to be designed from a premise of high risk. This includes aiming for short-term gains, sequencing longer-term reform agendas into discrete items with shorter time frames, and avoiding overburdened programs.

Box 4.2. The Coronavirus Pandemic and IFC Project Ratings

The coronavirus pandemic represents a shock to International Finance Corporation projects, putting the economic and financial sustainability of those projects at risk, at least temporarily. As one response, the International Finance Corporation and the Independent Evaluation Group are discussing how to adjust ratings processes and methodologies to account for shocks like the pandemic, which make the projects' implementation environment and country context more challenging. Proposals include making project objectives more realistic by rating projects based on their midcourse correction targets rather than those set at approval (before the shock occurred), and giving the International Finance Corporation more flexibility to choose the evaluation timing, which may help projects recover and meet targets later.

Source: Independent Evaluation Group.





The outcome orientation findings suggest a need to rethink the approach to collecting outcome evidence beyond the project level. The existing self-evaluation instruments and results measurement systems aggregate data from individual projects, but higher-level outcomes result from the interplay of different projects over time, something that none of the Bank Group's existing self-evaluation instruments capture. In line with past practice, this *RAP* does not make formal recommendations. However, IEG's forthcoming evaluation of country programs' outcome orientation discusses how using a wider set of methods and a focus on contribution rather than attribution could help support longer-term thinking and engagement on how the Bank Group contributes to important country-level outcomes. This includes more flexibly accounting for shocks and necessary program adjustments and better capturing the Bank Group's contribution to institutional change in countries (box 4.3).

It would be helpful to differentiate the purpose of collecting outcome evidence. At project level, setting objectives and assessing achievements that can be attributed to Bank Group support continues to be important for the institution's accountability and credibility. This requires realism when setting projects' development objectives as discussed in box 4.1. But beyond the project level, for results in country programs and at thematic levels, the purpose of collecting outcome evidence should not be to track and report and hold the institutions accountable for attributable results. Assessing outcomes often requires dedicated, context-specific evidence, which does not always lend itself easily to portfolio-wide aggregation. Outcome evidence can be robust when based on sound evaluation methods that use plausible theories of change and credible data to relate Bank Group activities to observed outcomes in sectors and countries.

Box 4.3. A Fresh Approach to Understanding Country Outcomes

The Independent Evaluation Group’s evaluation of country programs’ outcome orientation finds that a satisfactory self-evaluation instrument would need to go beyond the present approach, which is centered on “results frameworks premised on metrics, attribution, and time-boundedness” (World Bank 2020, x). A self-evaluation instrument suited for collecting higher-level outcome evidence would have to cover a longer period and focus on a sector or country to capture contributions to outcomes and assess the cumulative effects from multiple World Bank, International Finance Corporation, and Multilateral Investment Guarantee Agency lending, knowledge, and convening interventions. “A renewed country-level results system could conceive accountability differently, based on evidence of achievement and failures and description of learning and adaptation. It could acknowledge that the Bank Group can influence but not control country outcomes. It could recognize that country teams cannot decide all targets and objectives at design but must adapt during implementation” for reasons relating to shocks, uncertainty, changing circumstances, and, especially for the International Finance Corporation and the Multilateral Investment Guarantee Agency, unpredictable client demand. And it could realize that capturing contributions to country outcomes and assessing cumulative effects from multiple interventions requires dedicated evaluation inquiries, not just measurement of indicators. Data for such a renewed system could come from existing project evaluations, impact evaluations, ratings, stakeholder surveys, and other sources.

Source: World Bank 2020.

Looking Ahead

IEG plans to continue producing annual *RAPs* that aim to provide a broad perspective on the Bank Group’s performance. Though the exact shape of future *RAPs* is still undecided, IEG will continue its efforts to offer a lens through which to understand outcomes and outcome levels across sectors and Bank Group institutions. The Bank Group exists to work with its client countries on improving human conditions. A clear focus on outcomes helps it stay on course.



Bibliography

- Custer, Samantha, Zachary Rice, Takaaki Masaki, Rebecca Latourell, and Bradley Parks. 2015. *Listening to Leaders: Which Development Partners Do They Prefer and Why?* Williamsburg, VA: AidData.
- Denizer, Cevdet, Daniel Kaufmann, and Aart Kraay. 2011. “Good Countries or Good Projects? Macro and Micro Correlates of World Bank Project Performance.” *Journal of Development Economics* 105 (November): 288–302.
- Geli, Patricia, Aart Kraay, and Hoveida Nobakht. 2014. “Predicting World Bank Project Outcome Ratings.” Policy Research Working Paper 7001, World Bank, Washington, DC.
- Honig, Dan. 2018. *Navigation by Judgment: Why and When Top-Down Management of Foreign Aid Doesn’t Work*. Oxford: Oxford University Press.
- Honig, Dan. 2020. “Actually Navigating by Judgment: Towards a New Paradigm of Donor Accountability Where the Current System Doesn’t Work.” Policy Paper 169, Center for Global Development, Washington, DC.
- Ika, Lavagnon A. 2015. “Opening the Black Box of Project Management: Does World Bank Project Supervision Influence Project Impact?” *International Journal of Project Management* 33 (5): 1111–23.
- Moll, Peter G., Patricia Geli, and Pablo Saavedra. 2015. “Correlates of Success in World Bank Development Policy Lending.” Policy Research Working Paper 7181, World Bank, Washington, DC.
- Raimondo, Estelle. 2016. “What Difference Does Good Monitoring and Evaluation Make to World Bank Project Performance?” Policy Research Working Paper 7726, World Bank, Washington, DC.
- Ralston, Laura. 2014. “Success in Difficult Environments: A Portfolio Analysis of Fragile and Conflict-Affected States.” Policy Research Working Paper 7098, World Bank, Washington, DC.
- World Bank. 2010a. “Responding to Floods in West Africa: Lessons from Evaluation.” Independent Evaluation Group Note, World Bank, Washington, DC.
- World Bank. 2010b. “Response to Pakistan’s Floods: Evaluative Lessons and Opportunity.” Independent Evaluation Group Note, World Bank, Washington, DC.



- World Bank. 2012. *The World Bank Group's Response to the Global Economic Crisis, Phase II*. Independent Evaluation Group. Washington, DC: World Bank. http://ieg.worldbank.org/sites/default/files/Data/Evaluation/files/crisis2_full_report.pdf.
- World Bank. 2013. *World Bank Group Assistance to Low-Income Fragile and Conflict-Affected States*: Independent Evaluation Group. Washington, DC: World Bank. http://ieg.worldbank.org/sites/default/files/Data/reports/fcs_eval_0.pdf.
- World Bank. 2016a. *Behind the Mirror: A Report on the Self-Evaluation Systems of the World Bank Group*. Independent Evaluation Group. Washington, DC: World Bank. http://ieg.worldbank.org/sites/default/files/Data/Evaluation/files/behindthemirror_0716.pdf.
- World Bank. 2016b. *Results and Performance of the World Bank Group 2015*. Independent Evaluation Group. Washington, DC: World Bank. http://ieg.worldbank.org/sites/default/files/Data/Evaluation/files/rap15_fullreport.pdf.
- World Bank 2017. *Crisis Response and Resilience to Systemic Shocks: Lessons from IEG Evaluations*. Independent Evaluation Group. Washington, DC: World Bank. <http://ieg.worldbank.org/sites/default/files/Data/reports/building-resilience.pdf>.
- World Bank. 2018a. *Engaging Citizens for Better Development Results*: Independent Evaluation Group. Washington, DC: World Bank. http://ieg.worldbank.org/sites/default/files/Data/Evaluation/files/Engaging_Citizens_for_Better_Development_Results_FullReport.pdf.
- World Bank. 2018b. *Results and Performance of the World Bank Group 2017*. Independent Evaluation Group. Washington, DC: World Bank. <http://ieg.worldbankgroup.org/sites/default/files/Data/Evaluation/files/rap2017.pdf>.
- World Bank. 2019. *Results and Performance of the World Bank Group 2018*. Independent Evaluation Group. Washington, DC: World Bank. <https://ieg.worldbankgroup.org/sites/default/files/Data/Evaluation/files/rap2018.pdf>.
- World Bank. 2020. *The World Bank Group Outcome Orientation at the Country Level*. Independent Evaluation Group. Washington, DC: World Bank.



Photo Credits

- Page 2 *Sifting seeds in a field along Red River in Northern Vietnam*
QD-VN001 World Bank | Quy-Toan Do / World Bank | Vietnam
- Page 4 *Happy students at a school in Uganda, Africa. Students raising their hands.*
1784638013 | Boxed Lunch Productions | Uganda
- Page 5 *Kuala Lumpur is the capital city of Malaysia, landscape view over rice field plantation farming in morning sunrise*
131496056 | Szepei, from Switzerland | Malaysia
- Page 6 *ANSA-AW (Affiliated Network for Social Accountability- Arab World) was officially launched on March 14-15, 2012, in Rabat, Morocco. The event was organized by the World Bank in collaboration with CARE International (Egypt). ANSA-AW is a multi-stakeholder regional network comprised of CSOs, media, private sector and government representatives.*
Hoel_120313_DSC_7346 | Arne Hoel / World Bank
- Page 11 *Irrigation system watering a crop of soy beans at field.*
663246409 | Fotokostic, from Serbia
- Page 13 *Rossing Uranium Mine lies about 70km inland from Swakopmund close to the small town of Arandis. It was founded by amateur geologist Capt. Peter Louw in 1928 largely owned by the Rio Tinto Group (69% shares). In 2006 the mine produced about 7% of the world production of primary produced uranium. The mine is a long term supplier of uranium to the world nuclear power industry. Haul trucks being repaired and serviced at the workshop on the mine.*
JH-NA070906_0049 | John Hogg / World Bank
- Page 15 *Thermo-solar power plant. Ain Beni Mathar Integrated Combined Cycle Thermo-Solar Power Plant.*
DS-MA111 | Dana Smillie / World Bank
- Page 17 *Andean family taking their live stock to grazing pastures in the Andes, Peru, South America*
298577849 | Duncan Andison, from U.K. | Peru
- Page 18 *Winding desert road in Wadi Rum, Jordan*
115982191 | Boris Stroujko, from Switzerland | Jordan

- Page 21 *Soap and Water for Clean Hands for African Children*
1007424487 | Riccardo Mayer, from Germany
- Page 22 *Munnar, Kerala, India - October 12, 2007 : Beautiful landscape of a road passing a village and tree plantations near Munnar*
784009363 | ImagesofIndia | India
- page 23 *Baidoa / Somalia - March 2017 - People who carry water rest under a tree in the refugee camp.*
1100529911 | Mustafa Olgun, from Turkey | Somalia
- Page 24 *Low angle shot of modern glass buildings and green with clear sky background.*
613341923 | James Teoh Art
- Page 25 *The oil pump, industrial equipment*
1664994739 | Pan Demin, from China
- Page 26 *Kolony, Uganda – October 02, 2016: Many pregnant woman waiting for an ultrasound scan at the Kolonyi hospital in Uganda. A German doctor is there to educate the local doctors.*
766560016 | Dennis Wegewijs, from Germany | Uganda
- Page 29 *Metro Manila / Philippines - April 2019: Bonifacio Global city skyline at Magic hour. Bonifacio Global City or BGC, is a financial and lifestyle district in Metro Manila, Philippines.*
1367780063 | Hit Uno, from Japan | Philippines
- Page 30 *Daily life in Monrovia, Liberia on December 2, 1014.*
Liberia_Scene_Setters_0004 | Dominic Chavez/World Bank
- Page 33 *Daily life in Monrovia, Liberia on December 2, 1014.*
Liberia_Scene_Setters_0002 | Dominic Chavez/World Bank
- Page 34 *Indigenous peruvian Quechua woman with traditional hat and textile along a Andes road, Sacred Valley of the Inca, Cusco, Peru*
1802841382 | Sebastien Lecocq, from Belgium | Peru
- Page 35 *Workers building a new road.*
Albes Fusha / World Bank | Albania
- Page 36 *Good quality cocoa beans that are carefully selected in the hands of the owners of agricultural workers*
1389682223 | Attasit saentep, from Thailand

- Page 38 *According to the World Bank's Malaysia Economic Monitor, June 2013, the country's recent economic performance and near term outlook owes much to the commodities sector which includes palm oil. Palm oil is used for products such as animal feed.*
Nafise Motlaq / World Bank | Kuala Lumpur, Malaysia.
- Page 40 *Cattle and donkeys near a water point in Kenya's Eastern Province.*
FP-KE-0639 | Flore de Preneuf / World Bank
- Page 41 *Young boys on fishing boat.*
AH-GH061111_5002 | Arne Hoel / The World Bank | Ghana
- Page 42 *Cleaning solar panels. Ain Beni Mathar Integrated Combined Cycle Thermo-Solar Power Plant*
DS-MA117 | Dana Smillie / World Bank
- Page 43 *Local Intha woman weaving blue lotus fabric on a loom at the local lotus cloth weaving workshop at Inle lake, Shan State, Myanmar (Burma). January 2019. Selective focus*
1359398909 | Anya Newrcha, from Russia | Burma
- Page 44 *Abdul Satar, 30, says, before the cementing of the floor of the canal and the establishment of the sluices they had a lot of problems with the irrigation of their farms. 10/26/2014. Deh Surkh Village, Zenda jan district, Herat, Afghanistan.*
Ghulam Abbas Farzami / World Bank
- Page 46 *Zaheda feeding her chicken in the farm. Livestock Extension, FFS methodology training, National Horticulture and Livestock Project. 27 Jan 2015, Itifaq Mena, Surkhrud district, Jalalabad, Afghanistan.*
ABBAS Farzami / Rumi Consultancy / WorldBank
- Page 49 *Solar panels in desert under colorful sunset sky clouds, sun energy and electricity generation in Africa. Investment project to reduce greenhouse gas emissions.*
1384724600 | Yasmin Meraki, from Netherlands | Africa
- Page 52 *Water Projects, Lesotho. Advance Infrastructure of the Metolong Dam and Water Supply Programme included bridges (two) and a tarred access road of 32km road to the site from Maseru. Also power supply, water and sanitation, telecommunications, construction camp and permanent operational facilities. Bridge 1 over the Phuthiatsana River at Ha-Makhoathi. There is also small scale agriculture next to the river some of which is irrigated, Lesotho farmers however rely more on rainfall than irrigation.*
JH-LS-090625-2 | John Hogg / World Bank



- Page 54 *Parched soil by the White Nile.*
AH-SD2161869 | Arne Hoel / World Bank | Khartoum, Sudan.
- Page 55 *A Metrobus system bus, part of the new mass transportation system in Panama City, Panama.*
Gerardo Pesantez / World Bank
- Page 56 *African health professional or physician wearing face mask for protection and scrubs, child wearing wearing homemade mask sitting on her, looking at camera in covid-19 pandemic*
1798128712 | Yaw Niel, from Ghana
- Page 59 *Baobab trees along the rural road at sunny day*
187374941 | Dudarev Mikhail, from Russia | Senegal
- Page 61 *New rural roads have provided access to markets for the local communities.*
Ana Gjokutaj / World Bank | Albania





WORLD BANK GROUP
World Bank • IFC • MIGA