# Results and Performance of the World Bank Group 2021

IEG Independent Evaluation Group
WORLD BANK GROUP
World Bank • IFC • MIGA
WHAT WORKS

1818 H Street NW
Washington, DC 20433
Telephone: 202-473-1000
Internet: www.worldbank.org

**ATTRIBUTION**
Please cite the report as: World Bank. 2021. *Results and Performance of the World Bank Group 2021*. Independent Evaluation Group. Washington, DC: World Bank.

**COVER PHOTO**
Shutterstock/A_Lesik

**EDITING AND PRODUCTION**
Amanda O'Brien

**GRAPHIC DESIGN**
Luísa Ulhoa

# Results and Performance of the World Bank Group 2021

*December 23, 2021*

# Contents

## Boxes



## Figures

## Tables

## Appendixes

# Abbreviations

| | |
|---|---|
| AIMM | Anticipated Impact Measurement and Monitoring |
| COVID-19 | coronavirus pandemic |
| CY | calendar year |
| DPF | development policy financing |
| EvNote | Evaluation Note |
| FCS | fragile and conflict-affected situation |
| FY | fiscal year |
| GP | Global Practice |
| ICR | Implementation Completion and Results Report |
| ICRR | Implementation Completion and Results Report Review |
| IDA | International Development Association |
| IEG | Independent Evaluation Group |
| IFC | International Finance Corporation |
| M&E | monitoring and evaluation |
| MIGA | Multilateral Investment Guarantee Agency |
| MS+ | moderately satisfactory or above (for World Bank projects); mostly successful or better (for IFC investments and IFC advisory services) |
| PCR | Project Completion Report |
| PES | Project Evaluation Summary |
| RAP | *Results and Performance of the World Bank Group* (report series) |
| S+ | satisfactory or above (for MIGA and World Bank projects) |
| XPSR | Expanded Project Supervision Report |

*All dollar amounts are US dollars unless otherwise indicated.*

# Acknowledgments

This report was prepared by an Independent Evaluation Group (IEG) team led by Elena Bardasi, senior economist, under the overall direction of Alison Evans, Director-General, Evaluation, and with the guidance and supervision of Galina Sotirova, manager, Corporate and Human Development, and Oscar Calvo-Gonzalez, director, Human Development and Economic Management.

The core team included Ichiro Toda (who led the International Finance Corporation [IFC] and Multilateral Investment Guarantee Agency [MIGA] part), Jean-Jacques Ahouansou, Joy Behrens, Sylvie Bishweka, Kwabena Boasiako, Mariana Branco, Valentine Gandhi, Judith Gaubatz, Elisabeth Goller, Jessica Meckler, Chikako Miwa, Xiaoxiao Peng, Paola Rodriguez, Junko Sekine, Shiva Sharma, and Yi Yao. Maximillian Ashwill was the lead editor of the report.

Other IEG colleagues who made valuable contributions were Feruza Akbarovna Abduazimova, Harsh Anuj, Ana Belen Barbeito, Leonardo Alfonso Bravo, Unurjargal Demberel, Giuseppe Iarossi, Ramachandra Jammi, Jennifer Keller, Beata Lenard, Jose Candido Carbajo Martinez, Daniel Nogueira-Budny, and Stephan Wegner.

The technical advisory panel comprised Tamar Manuelyan Atinc (retired World Bank staff), Navin Girishankar (former IEG evaluator and World Bank and IFC staff), and Patricia Rogers (CEO, BetterEvaluation).

The team is grateful to the many IEG colleagues who shared their insight and feedback throughout the evaluation and to the Operations Policy and Country Services, IFC, and MIGA colleagues who engaged with us during the preparation of the report.

# Overview

The *Results and Performance of the World Bank Group 2021* (*RAP 2021*) report by the Independent Evaluation Group (IEG) reviews the World Bank Group's development effectiveness up to fiscal year (FY)20. The Bank Group includes the World Bank (comprising the International Bank for Reconstruction and Development and the International Development Association [IDA]), the International Finance Corporation (IFC), and the Multilateral Investment Guarantee Agency (MIGA).

The Bank Group's project-level ratings improved across the board according to the latest IEG-validated data. For the World Bank, the percentage of projects rated moderately satisfactory or above (MS+) in FY20 rose to 88 percent—a historic high. For IFC, the percentage of investments rated MS+ rose from 42 percent in calendar years (CY)16–18 to 47 percent in CY17–19. For MIGA, 68 percent of projects were rated satisfactory or better in the cohort FY14–19.

The higher ratings for the World Bank were not caused by disruptions to the self-evaluation and validation process from the coronavirus (COVID-19) pandemic. However, it is too early to tell whether COVID-19–related impacts on ratings will show more strongly in the future. That said, we identify several new factors not previously explored in other *RAP*s or IEG evaluations that can influence ratings. These factors include a project's novelty (defined as new or expanded elements in successor projects), selection of indicators and targets, outcome types, and outcome potential (box O.1). Overall, we find that Bank Group teams can improve performance by being innovative in project design and using strong measurement practices to track results.

### Box O.1. Methodology

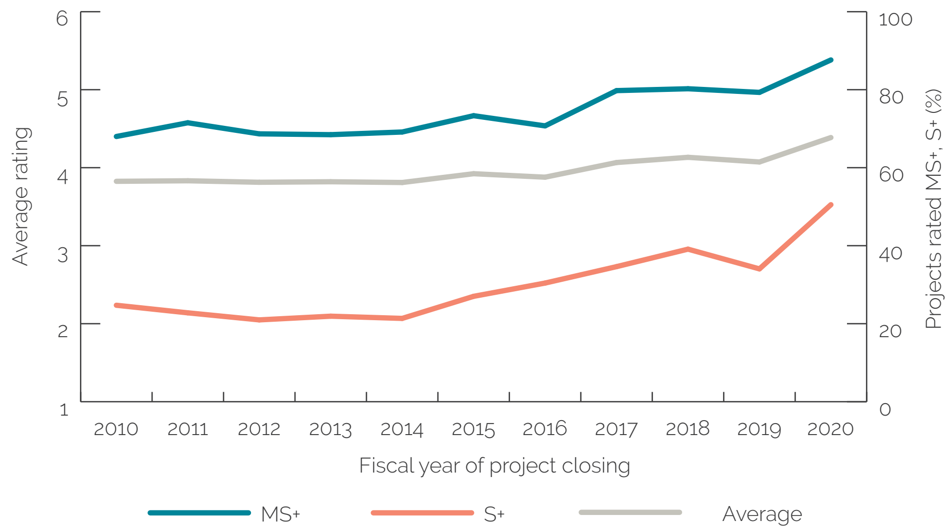
The *Results and Performance of the World Bank Group 2021* uses a novel methodology to expand on previous results and performance reports. First, we carry out an in-depth analysis of recent trends for both the World Bank and the International Finance Corporation (IFC). For the World Bank, we analyze a recent jump in project outcome ratings from fiscal years 2019 to 2020. For IFC, we analyze the uptick in ratings of investment projects during calendar year 2019 after several years of declining ratings and a reversal of the trend in calendar year 2018. Second, for the World Bank, we use matched data, linking successor projects in the Education and Transport Global Practices to their predecessor projects (in the same country and sector), to analyze the extent to which the World Bank either repeats project designs or introduces novelty to successor projects. We do this to detect signs of risk-averse or risk-taking behavior. Third, we analyze, in detail, the World Bank's selection of indicators and use of targets to understand how measurement practices affect ratings and performance. Fourth, for the World Bank, IFC, and Multilateral Investment Guarantee Agency, we look at the relationship between a project's outcome types and its results. For IFC, we also examine the relationship between a project's outcome potential and its ratings.

*Source:* Independent Evaluation Group.

## World Bank

The World Bank's project outcome ratings increased substantially in FY20 (figure O.1). This increase, which occurred for all categories of projects, extends the World Bank's positive ratings trend from the past several years and is the steepest of the past five years. Ratings increased for projects in all Practice Groups, and the increase was especially steep for Sustainable Development projects and for projects in West Africa and Europe and Central Asia. Ratings also improved in the most challenging places to operate, rising for projects in fragile and conflict-affected situations (FCS) and increasing in IDA countries. Ratings also increased notably for the World Bank's largest projects—those valued at over $100 million.

**Figure O.1.** World Bank Project Outcome Ratings, by Number of Projects




*Source:* Independent Evaluation Group.

*Note:* MS+ = moderately satisfactory or above; S+ = satisfactory or above.

Disruptions linked to COVID-19 did not appear to contribute to the jump in FY20 ratings or to bias ratings reporting. Higher ratings have been observed throughout the year, before and after the onset of COVID-19. A number of checks on the data confirmed that there was no apparent change in either the speed at which Implementation Completion and Results Reports and Implementation Completion and Results Report Reviews were processed or in the disconnect between their ratings. A check on the shares and ratings of investment policy financing and development policy financing did not find any evidence that the latest ratings increase was driven by changes in lending instruments.

Bank performance ratings, which include quality at entry ratings and quality of supervision ratings, also increased. For quality at entry, the ratings increase between FY19 and FY20 was substantial. Overall, many project categories that experienced large increases in project outcome ratings also experienced increases in Bank performance ratings: This was the case for projects in Sustainable Development, Europe and Central Asia, and IDA non-FCS countries and for large projects (those valued at over $100 million). For other project categories, such as projects in Western and Central Africa, project outcome ratings increased despite decreasing Bank performance. It is likely that higher Bank performance (and monitoring and evaluation [M&E]

quality—see following paragraph) has had a positive impact on the achievement of project outcomes; however, since all of these ratings are assigned at the same time (when the project outcome is already known), we cannot determine causality.

M&E quality ratings also increased, improving substantially between FY19 and FY20. A deeper analysis of M&E ratings shows that the robust increase in M&E quality had the strongest positive correlation with the recent increase in project-level efficacy ratings. Although correlation does not imply causation, these increases, in line with past RAPs and several studies, are indicative of improvements in the World Bank's ability to deliver better projects.

Notwithstanding general improvements in M&E quality, a closer look shows that higher project outcome ratings are not necessarily matched by higher quality indicators or more ambitious targets. The implication is that project teams and operational management are not systematically scrutinizing the selection of project targets and indicators, leaving an arbitrary space for deciding whether projects achieved their intended results or not. For instance, analysis shows that not all projects with institutional strengthening objectives have indicators to measure them. Many rely on weak, indirect, or anecdotal evidence with an overreliance on measured outputs over outcomes. Efficacy ratings for these projects were no worse, however, than the one-third of projects that did opt for a more direct and robust measurement approach. Although this is preliminary analysis, it could imply that final ratings are an imperfect indicator of whether or not intended outcomes are achieved and that there remain limited incentives in the system to adopt a more robust measurement approach.

There is a weak relationship between project efficacy ratings and the type of outcomes (that is, the type of intended change) a project aims to achieve. Our analysis indicates that only 4 outcome types out of 16 identified (expanded access to services, increased human capital, improved enterprise and sector performance, and enhanced equity and inclusion) have higher efficacy ratings than the others, but this is because they have higher M&E quality, which is what mostly drives the higher efficacy ratings.

One of the questions underlying the longer-term upward trend in ratings is whether it is a result of fewer risks being taken. This can express itself in a tendency to repeat project designs rather than embracing innovation. We looked at this in two Global Practices—Transport and Education—and found that successor projects that introduced novelty—introducing new or expanded elements over the previous project—performed as well as or better than projects that closely replicated the predecessor project. The results suggest that the World Bank has been able to take informed risks and introduce new elements relevant to each context without suffering lower project outcome ratings.

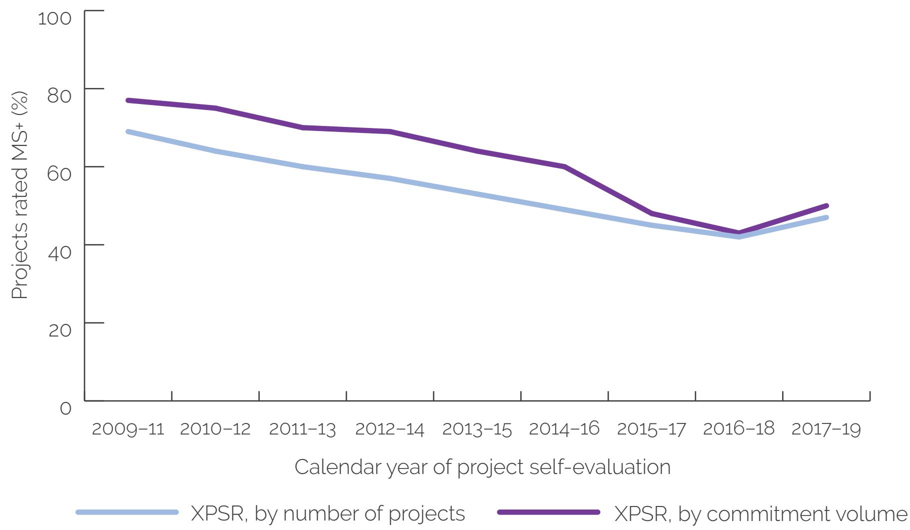
## International Finance Corporation

### Investment Projects

IFC's development outcome ratings for investment projects improved recently after several years of decline that continued until CY16–18 (figure O.2), when 42 percent of projects were rated mostly successful or better. When measured annually, the overall development outcome success rate was lowest in CY17, at 41 percent. However, in CY17–19, IFC's investment project ratings reversed this declining trend, improving to 47 percent mostly successful or better. The annual overall development outcome success rate improved by 18 percentage points to 60 percent in CY19. Three of IFC's four industry groups show a similar trend of improved ratings. This is good news, although it is too soon to conclude that the declining trend was completely reversed.

IFC's recent efforts to address negative influences on ratings may have paid off in the ratings improvements. Previous *RAP*s identified internal work quality issues, external risks, and broader market trends as factors that drive IFC's investment project ratings. In the past few years, IFC has created a new vice presidential unit to strengthen its project and macroeconomic analyses, launched the Anticipated Impact Measurement and Monitoring (AIMM) framework, and strengthened the Accountability and Decision-Making framework. IFC's management also improved the quality of self-evaluations. Although difficult to pinpoint precisely, it is likely that some of these efforts may be reflected in the recent ratings uptick. For example, after management's push to improve Expanded Project Supervision Reports (XPSRs), the share of XPSRs nominated as best practices increased from 12 percent in 2016, 11 percent in 2017, and 10 percent in 2018 to 20 percent in 2019. These efforts also increased the dialogue between IFC and IEG on project self-evaluations and reduced IEG-IFC ratings variance from 31 percent in CY17 to 8 percent in CY19.

Figure O.2. IFC Investment Project Development Outcome Ratings




*Source:* Independent Evaluation Group.

*Note:* IFC = International Finance Corporation; MS+ = mostly successful or better; XPSR = Expanded Project Supervision Report.

In CY19, IFC's subsector composition had fewer poorly performing clients and fewer greenfield projects (for the Financial Institutions Group), which also had a positive impact on ratings. In the Infrastructure industry group, there were fewer platform companies in the power sector, junior miners in the mining sector, and nonmobile telecom clients in the telecom, media, and technology sectors. These types of clients tend to have lower ratings than other clients. In Manufacturing, Agribusiness, and Services, the combination of fewer retail, tourism, construction, and real estate projects (whose performance declined in CY19) with more agribusiness and manufacturing projects (whose performance improved in CY19) contributed to IFC's aggregated improved ratings. For the Financial Institutions Group, the lower share of greenfield projects, which are projects that finance new ventures and activities and tend to have lower development outcome ratings, contributed to positive results in CY19 compared with previous years. Another factor behind the Financial Institutions Group's recently improved development outcome ratings is the improving ratings of projects in Europe and Central Asia, despite the Region's unstable economic environments.

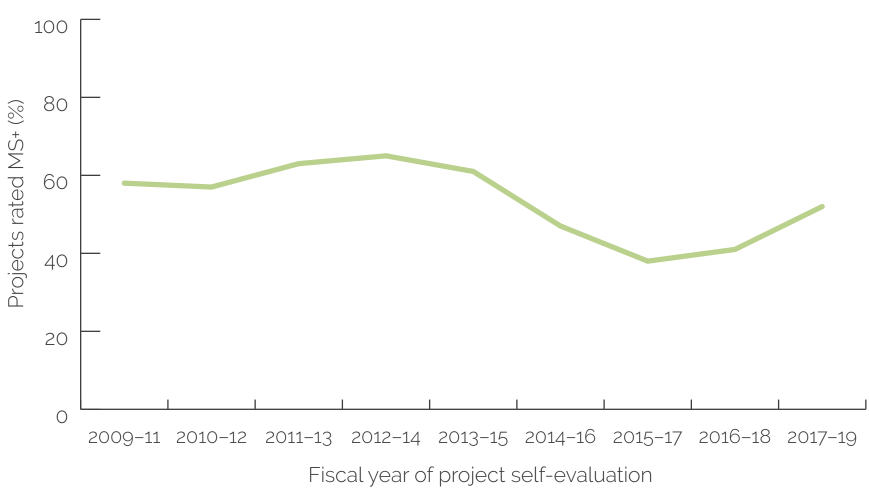
IFC projects are less likely to achieve market-level claims than project-level claims. Project-level claims, or project-level outcomes, are defined as a project's direct and indirect effect on stakeholders, the economy, and the environment. Market-level claims are derived effects, defined as a project's ability to catalyze systemic changes beyond those brought about by the project itself. Market-level outcome types also have a larger share of downgraded AIMM claim ratings than project-level outcome types. These results show that it is more difficult for IFC to achieve and measure market-level outcomes than project-level outcomes. Market-level outcomes depend on the broader market environment and external factors and are hard to attribute to IFC because individual projects generally have a minimal impact on the broader market. Market-level outcomes are also difficult to measure because they materialize over the long term and few indicators can measure a project's contributions with certainty. By contrast, project-level outcomes have shorter time horizons and often provide goods, services, financing, or infrastructure, all of which IFC and its counterparts have more control over achieving.

Projects with high development potential were not accompanied by lower XPSR ratings. A high development potential means a higher magnitude of development challenges in a given country and a more intense IFC contribution toward these challenges, as defined in IFC's AIMM framework. The fact that higher development potential did not lead to lower ratings undermines a common assumption that a higher development potential would contribute to lower ratings for IFC because of more sophisticated or challenging outcomes. Instead, the results show the opposite outcome: Projects with high development potential are not accompanied by lower XPSR ratings or higher variance in ratings. The results also show that IFC projects that addressed prominent corporate priorities—including climate change, IDA, FCS, and inclusive business (which includes gender)—do not have consistently lower ratings.

## Advisory Services

Development effectiveness ratings for IFC's advisory services projects continue to improve for several reasons. Development effectiveness ratings of mostly successful or better fell to their lowest level in FY15–17 but have been improving ever since (figure O.3). Previous *RAP*s show that several factors influence these ratings, including large project sizes, longer project durations, team leader changes, the client's commitment, IFC's work quality, and IFC's flexible and proactive supervision. IFC has taken actions to address these factors, possibly leading to better ratings. These actions have improved IFC's annual work quality ratings since FY18, particularly at project implementation and supervision. Moreover, IFC's Project Completion Reports have shown improved M&E and use of evidence, which likely contributed to improved development effectiveness ratings. For overall development effectiveness and outcomes, the share of Project Completion Reports that used quality evidence to a "sufficient extent" and "great extent" increased from 62 and 46 percent in 2016 to 70 percent for both categories in 2019. The improved evidence base may also have contributed to reducing the "variance gap" in Project Completion Reports, where the difference between IFC and IEG ratings decreased from 41 percent in 2016 to 13 percent in 2019.

**Figure O.3.** IFC Advisory Projects' Development Effectiveness Ratings, by Number of Projects



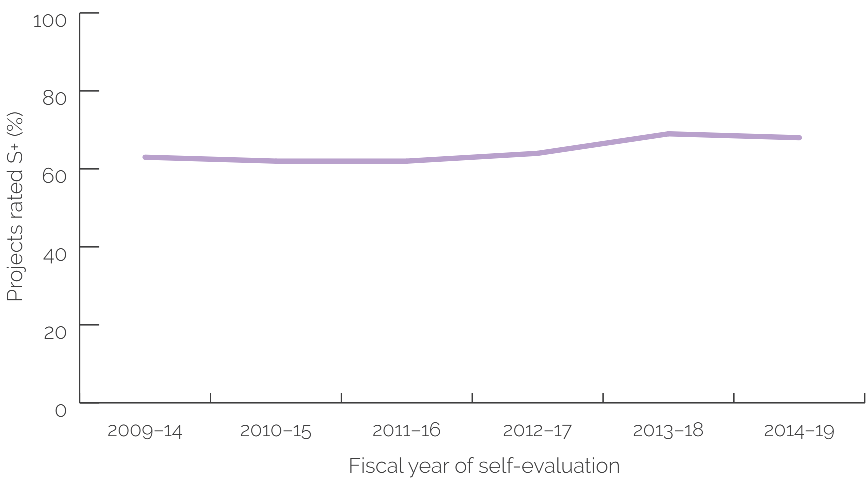

*Source:* Independent Evaluation Group.

*Note:* IFC = International Finance Corporation; MS+ = moderately satisfactory or above.

## Multilateral Investment Guarantee Agency

MIGA's project development outcome ratings have been increasing over the past 10 years. More specifically, MIGA's development outcome ratings increased from 62 percent satisfactory or above (S+) in FY11–16 to 68 percent S+ in FY14–19 (figure O.4). MIGA's financial sector had the lowest performance over this period, but its performance also improved. Among MIGA's four sectors, the Energy and Extractive Industries sector had the highest success rate of all the industry groups, although this has declined recently. MIGA's Agribusiness and General Services sector's performance was stable, and the Finance and Capital Markets sector's performance improved in 2018 and 2019.

Figure O.4. MIGA Project Development Outcome Ratings, by Number of Projects




*Source:* Independent Evaluation Group.

*Note:* MIGA = Multilateral Investment Guarantee Agency; S+ = satisfactory or above.

MIGA projects achieve project-level outcomes more often than foreign investment–level outcomes, and projects that address corporate priorities have mixed performances. From FY12–14 to FY17–19, project-level outcome achievement rates substantially increased. Meanwhile, foreign investment–level outcomes were less likely to be achieved. This is partially because of MIGA's inherent limitations, as a guarantee provider, in terms of collecting data on development results, particularly for foreign investment outcomes that rely on external factors for success. However, MIGA has made efforts to improve its self-evaluation. This suggests that MIGA's improved development outcome ratings are due to both increased evidence collection in recent years and improvement in performance. Meanwhile, projects addressing certain corporate priorities—including IDA, FCS, climate change, and South-South projects—did not experience a specific impact on ratings.

## Implications

- The World Bank and IEG could pay more attention to how well indicators measure project objectives. To do so would require a more systematic approach to gauging the appropriateness of indicators and targets early in the project cycle. A successful approach would include tightening the links between indicators and project objectives and defining targets in relation to scrutinized baselines.
- The World Bank could present ratings with more clarity about their strengths and limitations and could complement ratings with better information on the nature of the underlying development outcomes. IEG and the World Bank could periodically synthesize and report on development outcomes. Potentially, the World Bank could devise a system to regularly harvest project outcomes and key activities and match this information with ratings data for more integrated results and performance monitoring.
- IFC and MIGA could use information on outcome types and other characteristics to better assess risks, ratings, and development outcomes of projects. IFC's AIMM framework and MIGA's Impact Measurement and Project Assessment Comparison Tool framework already account for a project's estimated and actual development potential and development outcome risks. IFC and MIGA could take it a step further by assessing the prevalence of different outcome types and other characteristics in projects to help enhance the system.
- The Bank Group could further emphasize operating "on the frontier" as a goal in addition to meet the Corporate Scorecards rating targets. This shift in emphasis would encourage the Bank Group to inquire further about the motivations for risk taking, the evolution of project designs, the pursuit of corporate priority goals, and the best way to leverage internal resources and the client's

engagement, commitment, and capacity to deliver development results. This could help ensure that the Bank Group continues to selectively take risks to improve development outcomes.

# Management Response

Management of the World Bank Group institutions welcomes the Independent Evaluation Group (IEG) report, *The World Bank Group's Results and Performance 2021* (*RAP 2021*). Management welcomes IEG's positive overall findings on performance at the project level. The report's findings provide valuable insights to both learning and strategic decision-making.

## Overall

Management welcomes the overall positive findings, which place outcomes for projects at a historical high, and concurs with the overall conclusions. The report notes that 88 percent of Bank Group projects that closed in fiscal year (FY20) and were evaluated by IEG were rated moderately satisfactory or better at completion, surpassing corporate targets (75 percent), and underscores this impressive largest annual increase in outcome ratings over the past five years. These accomplishments "resulted from ratings improvements for virtually all categories of projects—all Practice Groups . . . all Regions . . . and almost all lending sizes" (72). Management is particularly pleased that "ratings also improved in the most challenging places to operate" (x) such as in fragile and conflict-affected situations (FCS) from 80 to 83 percent and International Development Association (IDA) countries from 75 to 86 percent. Management concurs with the *RAP 2021*'s overall conclusion that "longer-term ratings increases can occur with improved M&E [monitoring and evaluation] and selective risk taking derived from adding new activities in successor projects" (22). These variables correspond to management's past and current efforts, and the findings of this report will help further calibrate ongoing directions.

As suggested by IEG, management remains cautiously optimistic in interpreting these findings, as the effects of the coronavirus (COVID-19) pandemic are still unfolding. Although COVID-19 does not seem to have an immediate implication for the reviewed results, management is aware that projects closing in the next few years could see some negative effects brought through the crisis. Remote implementation support together with the inability of clients to access, collect, and validate data could also affect the strength of the evidence to support project effectiveness. Management looks forward to working with IEG to assess how best to interpret and rate development effectiveness in the years to come. Management commends IEG for improving the insightfulness of the *RAP* in 2021 and suggests more systematic dialogue to improve the clarity and shared understanding of its newly introduced methodological approaches. This is particularly true for the so-called novelty analysis. In addition, the introduction of outcome types as a follow up to the work on outcome levels in *RAP 2020* includes an interesting dimension that is currently being explored in the Bank Group's work on indirect pathways, as part of the outcome orientation road map. While noting that outcome types (as currently defined) do not have strong explanatory power with respect to the ratings, management sees value in continuing to refine the outcome typology. The typology currently presented would benefit from scrutiny from a larger group of experts from across the Bank Group. These conversations could be held within a reinvigorated Results Measurement and Evidence Stream that brings lessons and perspectives from across the Bank Group.

## Outcome Orientation

Management notes that this year's *RAP* departs from precedence by not assessing Bank Group performance in country engagements and regrets that the intersection of country and project results are not sufficiently clear in the report. Management recalls that the *RAP 2021* Approach Paper stated that "the *RAP 2021* analysis will be carried out at the project and country level. . . . At the country level, the focus will be on the association among aggregate project ratings across a country, patterns in outcome types across projects in a country, and country characteristics associated with greater challenge" (World Bank 2021, 8). As surfaced during the discussion on outcome orientation, including in the IEG evaluation *World Bank Group Outcome Orientation at the Country Level* (World Bank 2020d), the country level is the most relevant unit of analysis for outcomes, as the Bank Group supports country outcomes that are defined in clients' own strategies and reflected in country engagement products. Individual projects and their corresponding results frameworks should be viewed in the larger context of country engagements and wider strategies to aim for and monitor results. This is particularly true for understanding risk and results considerations, institutional strengthening results, and the desirability and feasibility of novelty. These foundational considerations for outcome orientation do not often occur in the vacuum of individual projects, but in the context of portfolio-wide efforts toward high-level results.

Management welcomes the report's conclusions regarding risk management in operations and notes its ongoing efforts to place outcome risk at the center of results programming, as part of the outcome orientation road map. *RAP 2021* concludes that the Bank Group "can discern when the conditions are right for projects to support novel and complex activities. In this way, teams can take informed risks and selectively build on past experiences to elevate a project's objectives without suffering lower project performance ratings" (75) confirming the findings of the outcome pilots, which showed that informed risk taking is an integral part of the model of how the Bank Group aims for outcomes. Management will reflect on the findings of *RAP 2021* as it continues rolling out three germane activities of the outcome orientation road map, namely (i) articulating a conceptual approach to assess development outcome risk in Bank Group programs and adjusting guidance, (ii) calibrating the Bank Group's Systematic Operational Risk Rating Tool, and (iii) building a more integrated focus on risk and results in training and outreach work on operations. These activities will help the Bank Group "inquire further about the motivations that drive risk taking [and] the evolution of project designs"(75) as suggested by the report.

Management welcomes the discussion on project novelty, yet it believes that the interplay between risks and results considerations at project design is significantly more complex and worth building on. Management appreciates the innovation in depicting risk, but notes some challenges in the definition of risk aversion and the new approach about "novelty." A significant part of the evaluation is focused on "risk aversion" and "novelty" as they relate to outcomes. A definition of risk aversion that is predicated on the inclusion of "new activities" with a differing "degree of difficulty" in successor projects is narrow and does not fully reflect the reality of choices that teams face in project design. It also does not reflect the interplay between project and country-level considerations. Experience has shown that project activities are chosen based on how appropriate they are in the context of client needs and how they are supported by analytical underpinnings, lessons learned, and stakeholder feedback, and not necessarily based on whether they are novel. Bank Group teams also periodically assess how risks could hinder the country's journey toward high-level outcomes through indirect and direct pathways and seek risk mitigation measures. Over time, investments in policy reform and institutional strengthening reduce the inherent capacity risks in each sector, so that the sector gradually moves from a higher to a lower capacity risk as institutions are strengthened to deliver effective services to achieve the relevant high-level outcome.

Assessing the desirability and feasibility of novelty requires a more nuanced understanding of how indirect pathways are followed and monitored, as management is doing as part of the Outcome Orientation Roadmap. Management welcomes the report's insights concerning institutional strengthening; these insights will inform the Bank Group's current exploration of possible methodologies to track indirect pathways to country-level outcomes to be incorporated in self-evaluation and reporting tools. Institutional strengthening activities often support longer-term, higher-level outcomes and are an essential step in helping clients achieve desired results. These activities should be viewed within the context of the entire causal chain leading to the intended outcome, as opposed to results in and of themselves. Better understanding institutional strengthening is essential to work "on the frontier" where real breakthroughs happen. Against the backdrop of renewed efforts to improve the strategic use of knowledge, management does not believe that the report has adequately reflected on how project analytical underpinnings inform the best combination of risks and opportunities and help determine the frontier at which Bank Group operations should work.

## Monitoring and Evaluation Quality

Management recognizes room for further improvements in the selection of indicators and targets, and it underscores that long-standing efforts are bearing fruit. The Bank Group has made significant progress in developing robust project-level monitoring and evaluation (M&E) systems, one of the foundations for delivering and tracking progress toward achieving country-level outcomes and instilling a culture of managing for outcomes. As noted by IEG, the recent introduction of theories of change, improvement in measurement frameworks, and the implementation of the new Implementation Completion and Results Report (ICR) methodology at the project level are already helping to make evident the connections between Project Development Objectives and high-level country outcomes. The Outcome Orientation Roadmap already includes several activities aimed at improving the way indicators are designed and utilized across key thematic priorities. In certain cases, management believes that some of the findings of the report concerning targets reflect strengths of the current self-evaluation system rather than shortcomings. This is the case for the 33 percent of targets that were "exactly achieved" (35). This not only shows that teams have become more realistic at calibrating expected results, but it also includes the types of indicators that one would expect to be exactly achieved (for example, number of assessments conducted, annual reports produced, new institutions created, and so on.). Previously, many teams set targets that were aspirational rather than necessary to achieve a development impact "breakeven" point and so were downgraded on failing to meet those aspirational targets. It is important to consider that the choice of outcome indicators could be limited in some instances by, for example, the availability, consistency, timeliness, and quality of data and therefore project teams would tend to choose indicators based on available data. These indicators quite often tend to be intermediate outcome rather than outcome indicators. Contrary to the report's conclusion, most teams and operational management have diligently focused on improving performance by specifically scrutinizing targets, therefore contributing to greater accuracy. Notwithstanding this progress, management recognizes room for further improvement, particularly in FCS and IDA FCS. In these cases, management has usually prioritized engagements to tackle the most pressing development challenges even when they were difficult to measure in such difficult contexts. The most recent practice note developed by the Bank Group on improving results in fragility, conflict, and violence (FCV) environments should support teams in improving the management of results in these countries (World Bank 2021a).

## Moving Forward

Management welcomes IEG's insight into specific project elements that relate to the achievement of project outcomes, particularly on the importance of M&E. Management is currently reviewing what has worked well concerning M&E quality to learn from different regions and Global Practices (the report does not disaggregate this rating) and to identify opportunities to generalize and scale-up good practices across the Bank Group. *RAP 2021* findings are timely and substantial to inform these ongoing efforts and management will involve IEG as this work progresses. As the report hints, many factors may have contributed to improved outcomes, with concerted actions to improve M&E as the most plausible: "These factors include teams improving measurement frameworks, preparing better theories of change, and becoming comfortable with the newer ICR methodology, among others. These improvements could have resulted from ICR reforms in 2017 or internal training and informal knowledge exchanges, which led to an increased focus by operational teams and development effectiveness units on building robust theories of change and paying more systematic attention to M&E quality. It is possible these factors came to fruition in FY20 and contributed to the ratings increase, but we could not measure whether this was the case" (21). Management firmly believes that any system that incentivizes appropriate quality will be grounded, not in one-time reviews, but in a well-constituted enabling environment. Regional Development Effectiveness units play a critical role in guiding project teams, with Operations Policy and Country Services support. Operations Policy and Country Services' learning arm, in cooperation with experienced operational staff, provides ongoing capacity-building to improve outcome orientation at the project and country levels. Processes such as the Quality Enhancement Review have supported project teams to better manage the ambitions and risks in projects. In short, improved project implementation, M&E, and outcomes result from a variety of staff roles and responsibilities, incentives, tools, systems, training, and resources.

Management believes that, to understand M&E quality appropriately, IEG and management should discuss and revisit the way this rating is measured. The M&E quality rating is derived from design, implementation and utilization, and further exploration could be helpful to build a deeper understanding in relation to the extent to which these ratings are Bank Group or client dependent, and if these ratings might be different at entry as opposed to at implementation or utilization. This may help assess where more effort should be placed in building capacities and providing support.

## International Finance Corporation Management Comments

Management of the International Finance Corporation (IFC) welcomes IEG's flagship report, *Results and Performance of the World Bank Group 2021*. For IFC, the report provides both an assessment of project performance and a review of the relationship between a project's outcome types and its results, and the relationship between a project's outcome potential and its ratings. Like the last edition, we welcome the specific focus on the World Bank, IFC, and Multilateral Investment Guarantee Agency (MIGA) with distinctive findings and suggestions for each organization because this helps each institution tailor its learnings and improvement initiatives. In addition, IFC management appreciates the exceptionally proactive engagement and collaboration IEG's *RAP 2021* task team extended to IFC throughout the process.

### International Finance Corporation's Development Outcomes and Effectiveness

Overall, IFC management is pleased with the improvements reported on both IFC's investment development outcome rating and advisory development effectiveness rating, as measured by the respective latest three-year cohorts. These improvements followed the reversals marked last year in the downward trends of the single-year-based ratings. In noting the uptick in both ratings, the report recognizes several initiatives that IFC has undertaken over the past few years to promote its focus on development impact: the launch of the Anticipated Impact Measurement and Monitoring (AIMM) framework, the establishment of Economics and Private Sector Development Vice Presidential Unit, IFC's greater attention to work quality, incentives and expert consultative resources for operational staff to focus on project evaluations, and the enhanced partnership and dialogue with IEG. IFC management welcomes IEG's acknowledgment of these significant efforts by IFC to substantiate the outcomes in project evaluations and engage with IEG in a constructive and substantial manner. As a result, there has been a sharp reduction in the ratings variance, and an enhancement in learning on both sides of IFC and IEG across investment and advisory services, as the report notes.

IFC management is cognizant that the full effects of the pandemic are yet to materialize. Over the coming years, we may observe a relative stabilization or even a decline of the ratings as the full effects of the pandemic unfold. Because the data and evidence—sometimes challenging to obtain even in normal times—may not be readily available, we hope that we can work with IEG more flexibly when we assess the impact of the projects that are affected by the pandemic. The conversation with IEG in this regard was initiated in FY20 and will continue.

Specifically, on the advisory side, IFC management welcomes IEG's recognition of the enhanced management attention to work quality and the use of improved evidence to monitor development results. Indeed, we have put considerable emphasis on ensuring that the Project Completion Reports have detailed evidence and that outcomes are well substantiated. In addition, teams have benefited from peer reviews within the results measurement team for borderline projects to ensure that the evidence is adequately presented to substantiate the outcomes.[1] As a result, the share of project completion reports that used quality evidence to a "sufficient extent" and "great extent" increased from 62 and 46 percent in 2016 to 70 percent for both categories in 2019, and we find it rewarding to be recognized for this. We hope to maintain this trend.

The report also highlights the lower development outcome ratings of IFC's investments in the Disruptive Technology and Funds sector. IFC management notes that the Disruptive Technology and Funds Expanded Project Supervision Report cohorts analyzed through the calendar year (CY)19 consisted almost entirely of equity projects. IFC's overall development outcome rating data by instrument show that IFC's evaluated stand-alone equity cohort has consistently performed below its loan and combined loan and equity cohorts and 17 to 19 percentage points below stand-alone loans in development outcomes in CY16–19. In fact, inadequate financial and development results of IFC's equity investments post-2009 prompted IFC management to adjust its equity approach in 2018.[2] The revised approach—

which focuses greater attention on macro and sector dynamics, the ability to meaningfully influence investee behavior, and the active management of key investments—has been under implementation for the past three years, and initial results are promising. However, it is too early for Expanded Project Supervision Report sampling to meaningfully analyze the impact of IFC's actions on turning around its equity performance.

## On the Special Analyses of International Finance Corporation Investment Services

IFC management welcomes IEG's innovative analyses investigating the relationships between a project outcome type and its results as well as between a project outcome potential and its ratings. We look forward to continuing the discussion with IEG colleagues on how we may best apply the analyses to our operational work and decision-making for improving performance. We also welcome IEG's efforts to integrate the AIMM framework into the analyses and reiterate the need for the results from the analyses of outcome type and outcome potential to be interpreted cautiously given the early stage of the AIMM monitoring pilot and the limited data set of AIMM assessments (particularly with respect to AIMM scores "backfilled" on investment projects in IFC's portfolio).

The report makes a good point about the challenges of achieving market-level claims. Given the factors at play in accelerating market development, it can be difficult for a single institution or project to catalyze such development. Nonetheless, IFC's ambition to "create (and develop) markets" remains a central pillar of its strategy and a strong focus of its investment and advisory work. That is why recent initiatives, including with respect to IFC's upstream engagement and deployment of scaled approaches, such as frameworks and platforms,[3] have been a management priority. Moreover, the AIMM framework deepens IFC's capacity to understand the extent of its ability to generate market changes and capture such changes in the monitoring and measurement of its projects. This is a welcome development, even if it means that actual results may not be as ambitious as originally expected, especially in the short- to medium-term. IFC is reviewing the periodicity and practicality of when and how best it would be able to fully capture market outcomes and has plans to prepare guidance for this shortly.

IFC management welcomes the preliminary conclusion that projects with high development potential and those supporting prominent corporate priorities are not associated with lower Expanded Project Supervision Report ratings. We have also confirmed this relationship on a preliminary basis using AIMM score data and have gained two important take-aways: (i) IFC should not shy away from undertaking projects in the most challenging environments—similar to one of the report's concluded implications that the Bank Group could further emphasize operating "on the frontier," and (ii) the separation of impact potential from likelihood of achievement continues to provide an important analytical construct for the assessment of development impact.

The report suggests that IFC consider assessing the prevalence of different outcome types and other characteristics in projects to help enhance the system. The practical implication of this suggestion is to incorporate risk (or likelihood of achievement) in the assessment of specific development impact claims. An important design feature of the AIMM framework is the use of a likelihood rating assigned at the dimension level (project outcomes and market creation). IEG's suggestion to consider incorporating the assessment of likelihood (or risk) at the claim level appears to be supported by the current analysis. At the same time, the nuances of such relationships merit further investigation given the analytical and workload implications for economists and project teams as well as the range of considerations that weigh on a project's AIMM rating during supervision. The work undertaken by IEG for this report makes an important contribution to IFC's analysis, which we hope to build on through our ongoing work related to inventorying AIMM claims.

## Microproduct Reform and Moving Forward

Finally, we are about to embark on a historic joint effort with IEG colleagues to redesign IFC's evaluation framework for investment projects—an opportunity created by IEG's microproduct reform. IFC management and teams look forward to constructive and focused partnership with IEG in this exercise, drawing on experience gained under the current system, while addressing the dual objective of accountability and learning. We appreciate IEG's keen interest in better aligning the new framework with the AIMM methodology through the redesign. IFC remains committed to focusing on development impact and looks forward to working together with IEG colleagues on the microproduct reform.

In the interim, IFC management hopes to continue exploring options with IEG for addressing known weaknesses of the existing evaluation framework by allowing some elements of the reform to be already applied to the current methodology and process as much as practicable. In this regard, we appreciate continuing dialogues on such efforts as finding practical methodology for benchmarking project performance and a sampling method to generate a cohort that more closely represents key aspects of IFC's portfolio. IFC management believes that these parallel resolutions will enhance the value of the upcoming annual evaluation programs as it looks forward to collaborating on the microproduct reform for a longer-term framework.

## Multilateral Investment Guarantee Agency Management Comments

The Multilateral Investment Guarantee Agency welcomes the *RAP 2021* report, and we are grateful to IEG for their engagement and dialogue on MIGA's data and analysis of the report.

### Multilateral Investment Guarantee Agency Performance

MIGA appreciates IEG's recognition of MIGA's development outcome ratings of satisfactory or better remaining near peak levels. MIGA's development outcome ratings increased from 62 percent in FY11–16 to 68 percent in FY14–19 by project number and from 59 percent to 74 percent by gross issuance amount across the same periods. Among MIGA's four sectors, the Energy and Extractive Industry sector had the highest development outcome ratings by project number at 75 percent. IEG recognized that MIGA supported several countries' "first-of-a-kind" power projects with strong potential to achieve demonstration effects.

*RAP 2021* illustrates that projects aligned with MIGA's corporate priorities had a clear tendency for better performance. Projects in IDA countries and FCS countries outperformed those not in these countries. These IEG findings give strong assurance of MIGA's ability to expand in these two priority areas, given that in FY21, a quarter of our guarantees supported projects in IDA countries and fragile situations. The *RAP 2021* analysis of the development impact of climate change projects was inconclusive, as IEG identified a small number of evaluated projects as addressing climate change. Since that period, MIGA has scaled up its climate finance activities, including enhancing and systematizing practices about identification of projects addressing climate change. In FY21, MIGA contributed to the Bank Group's second Climate Change Action Plan (2021–25), with 73 percent of MIGA projects supporting climate change mitigation or adaptation or both.

*RAP 2021* also connected better actual outcomes with a strong focus on results measurement. IEG concluded that MIGA's improved development outcome ratings are due to increased evidence collection in recent years and a definite performance improvement. *RAP 2021* acknowledged improved evidence collection due to MIGA's ongoing work to enhance the quality of our self-evaluation and evidence collection. Further to our emphasis on evaluation, MIGA's learning initiatives are another vital contributor to the strong performance. MIGA is continuing to conduct joint learning events with IEG with active participation by the self-evaluation teams. These joint learning events are targeted to reach all staff within MIGA and have been recognized as one of the good practices for learning from project self-evaluations. Moreover, MIGA has also successfully launched its ex ante development impact assessment tool, the Impact Measurement and Project Assessment Comparison Tool (IMPACT), which is serving to enhance focus on project selection and development impact by assessing potential project outcomes and foreign investment effects, combined with a likelihood factor for realizing them. MIGA believes attention to the results, both ex ante and ex post project approval, as well as learning from the self-evaluation process, including IEG's validation assessments, will contribute to "smarter" risk taking where new ideas and approaches may be needed to address increasing challenges for achieving results in MIGA's priority areas in IDA FCS, climate change, and recovery from the pandemic.

## Outcome-Type Analysis and Classifications

MIGA welcomes IEG's outcome-type analysis to examine the relationship between a project's outcome types and its results. Unlike IEG's analysis for IFC's investment projects, IEG's analysis was, for MIGA-evaluated projects, before the IMPACT framework was introduced. MIGA welcomes the *RAP 2021* outcome-type analysis to fill this gap; however, as IEG highlights, this exercise comes with many methodological challenges and is not equivalent to a full IMPACT assessment. Therefore, MIGA fully agrees with the importance of heeding the *RAP 2021* call for a cautious interpretation of the results of MIGA's outcome-type analysis.

IEG's analysis pointed out that MIGA projects have a higher probability of achieving project-level outcomes than foreign investment–level outcomes. Given the nature of systemic "beyond the project" effects of the foreign investment mobilized by MIGA guarantees, the IMPACT framework sharpens the ex ante assessment by critically assessing plausible changes the project can bring and the likelihood of realizing such effects. MIGA is looking forward to further dialogue with IEG on pragmatic ways of evaluating systemic foreign investment-level outcomes, the timing of achieving these effects, and credible and practical indicators that measure the changes.

## Moving Forward

The report suggests that MIGA consider assessing the prevalence of different outcome types and other characteristics in projects to help enhance the system. The full implementation of the IMPACT framework is intended to sharpen the focus of assessing the project's outcome types. Also, the system uses likelihood ratings for both project outcome and foreign investment effects to incorporate the risk for realizing development impact claims. MIGA would like to monitor IFC's experience of the AIMM framework and its likelihood adjustments, including costs and benefits of incorporating the assessment of likelihood (or risk) at the claim level, should IFC proceed to incorporate this change, and implications of likelihood adjustments for ex post evaluative development outcome ratings.

The report also suggests the Bank Group could further emphasize operating on the frontier as a goal in addition to meeting the corporate scorecards rating targets. MIGA added new impetus to its strategic focus on product innovation and product application that embraces the spirit of risk taking on the frontier and helps address the decline in foreign direct investment exacerbated further by COVID-19. Examples in FY21 include (i) new approaches to developing the solicitation and filtering of innovative ideas from staff and clients, (ii) progress on new approaches to scaling up climate finance activities, (iii) new solutions to MIGA's support for capital markets transactions, (iv) scaling up of capital relief approaches for financial institutions, and (v) the introduction of a new trade finance product in partnership with IFC. MIGA welcomes IEG's suggestions and looks forward to further dialogue on how operations on the frontier can be evaluated ex post without disincentivizing risk taking and innovative approaches.

[1] Borderline projects are those in the "gray zone" between mostly successful and mostly unsuccessful ratings on the development outcomes and development effectiveness rating scale. The changes between mostly successful and mostly unsuccessful take much greater significance in the International Finance Corporation's development outcome and development effectiveness scores than those between any other ratings in the current framework.

[2] "A New Approach to Investing in Equity," informal International Finance Corporation Board Meeting, November 29, 2018.

[3] Examples of new platforms include the Start-Up Catalyst, Global Trade Finance Program, Small Loan Guarantee Program, Small and Medium-Sized Enterprise Ventures, Fast-Track COVID-19 Facility, Base of the Pyramid, and Global Health Platform. For more information, see Strategy and Business Outlook Update FY22–24.

# 1. Introduction

The *Results and Performance of the World Bank Group* (*RAP*) *2021* report by the Independent Evaluation Group (IEG) reviews the World Bank Group's development effectiveness for fiscal year (FY)21. The Bank Group includes the World Bank (which comprises the International Bank for Reconstruction and Development and the International Development Association [IDA]), the International Finance Corporation (IFC), and the Multilateral Investment Guarantee Agency (MIGA). The *RAP*'s overarching question is, What does the existing evidence from IEG's project evaluation and validation work show about the Bank Group's results and performance?[1] The analysis to answer this question focuses on the Bank Group's project development outcomes, project outcome ratings, project efficacy ratings, and other project- and country-level characteristics. Building on previous *RAP*s, this year's approach provides a more comprehensive analysis of results and performance and assesses whether better ratings are indicative of better development results.

This *RAP* finds that the Bank Group's ratings improved for the World Bank, IFC, and MIGA over the past year. It shows that these recent improvements—particularly the jump in the World Bank's project outcome ratings in FY20—were driven by improvements across the board for all Practice Groups and Regions. The World Bank's ratings improvements were not caused by disruptions from the coronavirus (COVID-19) pandemic. For IFC, changes in its portfolio composition help explain the recent increase in ratings. For the World Bank, IFC, and MIGA, the ratings increases are good news, although it is too early to tell if these improvements will be sustained. The *RAP* identifies several factors not previously explored in other *RAP*s or IEG evaluations that can influence ratings. These factors include a project's level of novelty (defined as World Bank teams introducing new or expanded elements to successor projects in a sector and country), a project's selection of indicators and targets, its outcome types, and its outcome potential.

## Methodology

*RAP 2021* uses a novel methodology to expand on previous *RAP*s (see box 1.1 for key terms and concepts). First, it carries out an in-depth analysis of recent trends for both the World Bank and IFC. For the World Bank, it analyzes a recent jump in project outcome ratings from FY19 to FY20. For IFC, it analyzes the uptick in ratings during calendar year (CY)19 after several years of declining ratings and a reversal of the trend in CY18. Second, for the World Bank, the *RAP* uses matched data, linking successor projects in the education and transport sectors to their predecessor projects (in the same country and sector) to analyze the extent to which the World Bank either repeats project designs or introduces novelty to successor projects. The *RAP* does this to detect signs of risk-averse or risk-taking behavior. Third, the *RAP* analyzes, in detail, the World Bank's selection of indicators and use of targets to understand how measurement practices affect ratings and performance. Fourth, for the World Bank, IFC, and MIGA, the *RAP* looks at the relationship between a project's outcome types and its results. For IFC, the *RAP* also examines the relationship between a project's outcome potential and its ratings. Fifth, for both IFC and MIGA, the *RAP* analyzes whether IFC and MIGA projects addressing corporate priorities have higher or lower overall ratings. It should be emphasized that evaluation and rating approaches are different across the Bank Group organizations. As such, overall results and performance cannot be compared across them (for more information on evaluation approaches, see appendix A).

For the World Bank analysis, we used different groupings and samples of projects for different types of analyses. Chapter 2's analysis of the overall rating trends includes all projects that closed between FY00 and FY20 and had an Implementation Completion and Results Report (ICR) and ICR Review (ICRR) completed by August 10, 2021 ($N$ = 5,825), with a special focus on the most recent period, FY10–20 ($N$ = 3,080).[2] The in-depth analysis of novelty, indicators, and outcome types uses a sample of projects that closed in FY12–14 and from FY17 to the second quarter of FY20 and are representative at the level of period and Practice Group (Sustainable Development, Infrastructure, Human Development, and Equitable Growth, Finance, and Institutions).[3] The samples are 90 percent representative with a 10 percent margin of error. Appendix A shows the size and composition of this sample. For the most part, we relied on ICRRs as the key source of information; however, specific parts of the analysis are also based on information derived from Project Appraisal Documents and ICRs. Appendix A also presents other samples and sources of information that the team used in the analyses.

### Box 1.1. Key Terms and Concepts

**Outcome:** A change in behaviors, conditions, or situations resulting from World Bank Group activities. Outcomes may include intended, unintended, positive, and negative changes.

**Theory of change:** The logic, expressed in project design documents and the Implementation Completion and Results Report, that identifies expected cause-and-effect relationships among inputs, activities, outputs, intermediate results, outcomes, impacts, and underlying critical assumptions.

**Project development objective (PDO):** A World Bank project's stated objective, framed as a positive outcome.

**Project claims and market and foreign investment claims:** The Independent Evaluation Group considers these equivalent to the World Bank's PDO in the Anticipated Impact Measurement and Monitoring system of the International Finance Corporation (IFC) and the Impact Measurement and Project Assessment Comparison Tool system of the Multilateral Investment Guarantee Agency (MIGA).

**Self-evaluation:** A formal assessment of a project, program, or policy conducted by or for those in charge of the activity. In the Bank Group, self-evaluation takes the form of a systematic written account of the results and performance of a project or operation, and those in charge assign ratings based on criteria defined in guidelines to assure comparability among reports.

**Validation:** The Independent Evaluation Group's independent, critical review of the evidence, results, assessments, and ratings from self-evaluation.

**Ratings:** Ratings are rubrics for assessing performance relative to a project or program's objectives. Ratings summarize the self-evaluation narrative into categories or values that enable aggregation across operations. Examples include the following:

*(continued)*

## Box 1.1. Key Terms and Concepts (cont.)

- For World Bank projects, the "outcome" rating brings together three underlying dimensions: relevance, efficacy (achievement of objectives), and efficiency. Independent Evaluation Group validations assign ratings for a project's efficacy in achieving each of its individual objectives and for overall efficacy in achieving the project development objective. Other key ratings are quality at entry (which, together with the quality of supervision rating, determines the Bank performance rating) and monitoring and evaluation quality.
- For IFC investment projects, the "development outcome" rating brings together four underlying dimensions: project business success, economic sustainability, environmental and social effects, and private sector development. IFC's development outcome ratings do focus on the achievement of expected objectives and a project or company's results against several benchmarks (such as the performance of peers, the market, or similar industries) and unintended outcomes (positive and negative). Other key ratings are for IFC's additionality, IFC's investment outcomes, and IFC's work quality.
- For IFC advisory services projects, the "development effectiveness" rating brings together five underlying dimensions: strategic relevance, outputs, outcomes, impacts, and efficiency. Other key ratings are IFC's work quality and IFC's role and contribution.
- For MIGA projects, the "development outcome" rating brings together four underlying dimensions: project business success, economic sustainability, environmental and social effects, and foreign investment effects. Other key ratings are MIGA's strategic relevance, MIGA's role and contribution, and MIGA's assessment, underwriting, and monitoring, which are aggregated under MIGA's effectiveness ratings.

Monitoring and evaluation (M&E) quality: For World Bank projects, M&E quality is assessed at the project level and comprises M&E design, implementation, and use.

- M&E design is assessed based on the extent to which (i) the theory of change was sound and reflected in the results framework; (ii) the objectives were clearly specified; (iii) the indicators encompassed all outcomes of the PDO statement; (iv) the intermediate results indicators could adequately capture the contribution of the operation's components (activities) and outputs toward achieving PDO-level outcomes; (v) the indicators were specific, measurable, achievable, relevant, time-

*(continued)*

## Box 1.1. Key Terms and Concepts (cont.)

bound, and had baselines and targets available; (vi) the measurement methods were adequate; and (vii) the arrangements were well embedded institutionally.

- M&E implementation is assessed based on the extent to which (i) planned baseline data collection was carried out; (ii) the indicators included in the results framework were measured and reported; (iii) weaknesses (if any) in M&E design—including specification of indicators—were corrected during implementation; (iv) the agency responsible for M&E (and any other relevant stakeholders) ensured attention to effective M&E implementation; (v) data used for M&E were found to be reliable and of good quality (important elements here include sound methodology, independence of analysts, and quality control); (vi) if relevant, beneficiaries were involved in defining target indicators and assessing their achievement; and (vii) M&E functions and processes are likely to be sustained after project closing.
- M&E use is assessed based on the extent to which (i) M&E findings were communicated to the various stakeholders (for example, to inform adaptive management); (ii) M&E information led to strategic redirection or resource allocation or to other positive or negative shifts in the implementation of the project or program; (iii) M&E data were used to provide evidence of achievement of outcomes and not just to provide evidence of application of inputs or achievement of outputs; and (iv) M&E data or findings have informed subsequent interventions or are expected to influence subsequent interventions in the near term.
- High M&E quality helps clarify the "line of sight" from projects to high-level country outcomes by explicitly defining and demonstrating project-level outcomes.

*Source:* Independent Evaluation Group.

For IFC's analysis, the sample includes projects with evaluations that IEG validated from CY09 to CY19 for investment projects and FY09 to FY19 for advisory services.[4] For the outcome type analysis on investment projects, the sample includes (i) projects evaluated by IFC and validated by IEG during FY12 and the first half of FY20 (evaluation cycles), and (ii) projects with backfilled Anticipated Impact Measurement and Monitoring (AIMM) data.[5] The analysis compared the early period (CY12–16) and the later period (CY17–20). We relied on Expanded Project Supervision Report (XPSR), Evaluation Note (EvNote), and AIMM data as the main sources of information.

For MIGA's analysis, the sample includes all projects with evaluations validated by IEG during FY09–19. For the analyses of outcome types, the sample includes all projects validated by IEG by December 2020 for the FY12–14 and FY17–19 cohorts. We relied on Project Evaluation Reports and Validation Notes as the main sources of information.

[1] The *Results and Performance of the World Bank Group (RAP)* reports measure performance through ratings assigned during the Implementation Completion and Results Report Review (ICRR) validation of the Implementation Completion and Results Report (ICR) project self-evaluation.

[2] Ratings for World Bank projects in this report are aggregated according to the closing fiscal year of the project, so fiscal year (FY)20 here refers to the group of World Bank projects that closed in FY20, completed their ICRs, and had their ICRs validated by the Independent Evaluation Group (IEG) in ICRRs by August 10, 2021.

[3] The more detailed definition of the FY17 cohort was projects closed in FY17, FY18, FY19, or the first two quarters of FY20 and that had ICRRs completed with ratings available in the system as of January 21, 2021. For additional information, see appendix A.

[4] For the self-evaluation reports of International Finance Corporation (IFC) investment projects (Expanded Project Supervision Report) and advisory projects (Project Completion Report) as well as Multilateral Investment Guarantee Agency projects (Project Evaluation Report), the trend data reported in this *RAP* includes those self-evaluation reports validated by IEG by August 10, 2021.

[5] In the backfilling exercise, the IFC retroactively applied the Anticipated Impact Measurement and Monitoring (AIMM) framework to projects whose approval predated the AIMM framework, including identification of outcome claims, conducting underlying analysis of development outcomes, indicator targets and results, and assignment of corresponding ratings (collectively we call these AIMM data in this *RAP*). This exercise identified outcome claims and their expected results and verified the results for each outcome claim. The IFC projects to which AIMM was applied at approval have not yet been evaluated by IEG.

# 2. World Bank Results and Performance

The World Bank's project outcome ratings increased substantially in FY20. This increase, which occurred for all categories of projects, extends the World Bank's positive ratings trend from the past several years. The analysis found that these improved ratings are compatible with selective risk taking that derives from introducing novelty into projects. This chapter also finds that ratings increases are consistent with improvements to monitoring and evaluation (M&E) quality, although the World Bank's outcome indicators and targets do not always robustly measure development results.

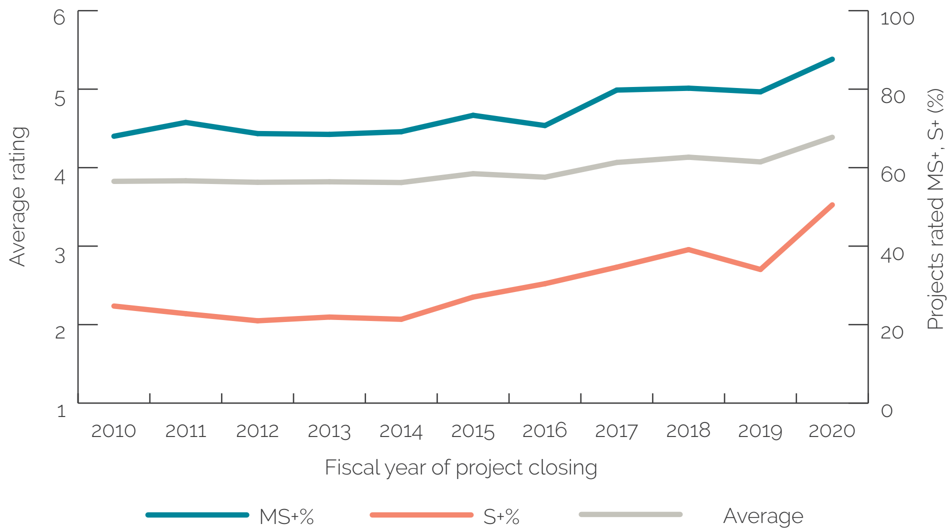
## The Fiscal Year 2020 Ratings Increase

The World Bank saw a recent jump in project outcome ratings for all project categories in FY20; if confirmed,[1] this increase would be the steepest of the past five years. This section presents trends and explores several factors associated with World Bank project performance,[2] as identified in previous *RAP*s and the broader literature, such as a project's size, lending group, fragile and conflict-affected situation (FCS) status, quality at entry, and M&E quality. Some of these factors correlate with the positive change in ratings but do not explain the extraordinary jump in FY20.

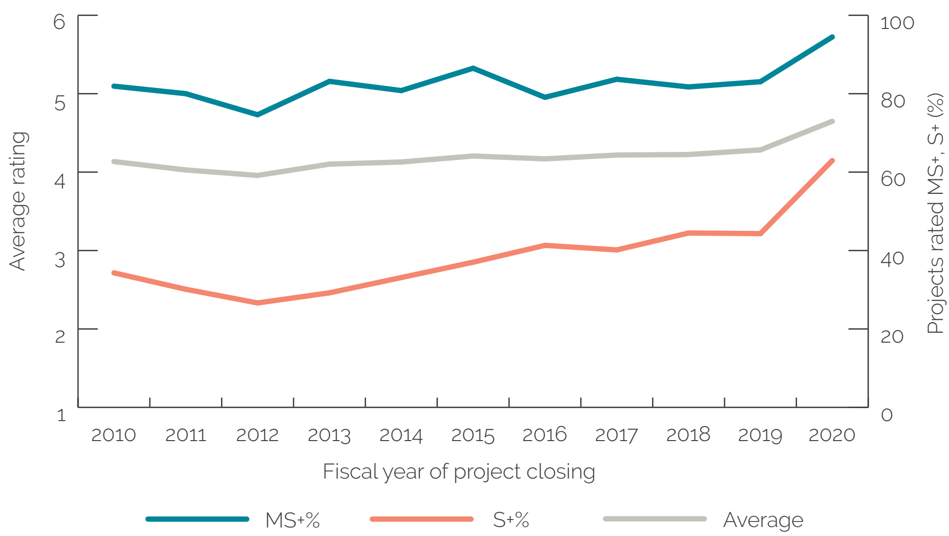
The World Bank's project outcome ratings improved dramatically from FY19 to FY20. This jump in ratings occurred for both average project outcome ratings, which increased from 4.1 to 4.4 on a scale of 1 to 6 (figure 2.1, panel a, gray line), and percentage of projects with outcome ratings of moderately satisfactory or above (MS+), which increased from 79 percent to 88 percent (figure 2.1, panel a, blue line). This annual increase was the largest over the past five years.[3] The increase was similar whether measured by the volume of projects (figure 2.1, panel b) or by the number of projects (figure 2.1, panel a).

**Figure 2.1.** World Bank Project Outcome Ratings

a. By number of projects




b. By volume of projects

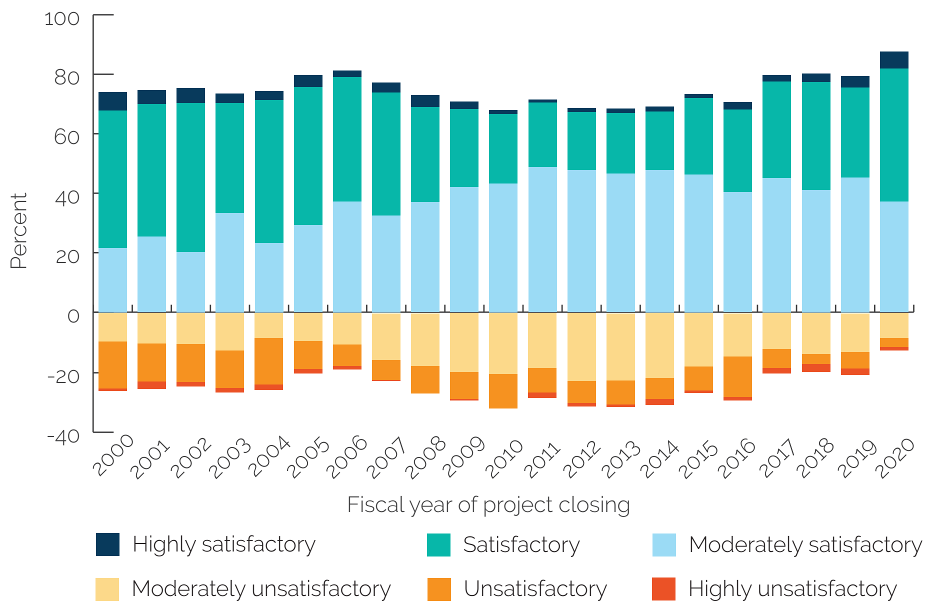



*Source:* Independent Evaluation Group.

*Note:* The left-hand axis represents the ratings scale for outcome ratings of World Bank projects, in which 1 represents a rating of highly unsatisfactory, 2 unsatisfactory, 3 moderately unsatisfactory, 4 moderately satisfactory, 5 satisfactory, and 6 highly satisfactory. The right-hand axis represents the percentage of projects with outcomes rated MS+ or S+. MS+ = moderately satisfactory or above; S+ = satisfactory or above.

The increase in the percentage of projects rated satisfactory and highly satisfactory was especially steep. The increase in the percentage of projects whose outcome ratings were satisfactory or above (S+; figure 2.1, panel a, red line) climbed from 34 percent in FY19 to 51 percent in FY20. Nearly 6 percent of projects that closed in FY20, with ICRs submitted and validated by IEG, were rated highly satisfactory, the highest since 2001. Moreover, only 1 percent of FY20 projects were rated highly unsatisfactory, and 3 percent were rated unsatisfactory. Figure 2.2 shows that, essentially, the bottom tail in the ratings distribution is disappearing, whereas the top tail is thickening.

Figure 2.2. Distribution of the World Bank's Project Outcome Ratings

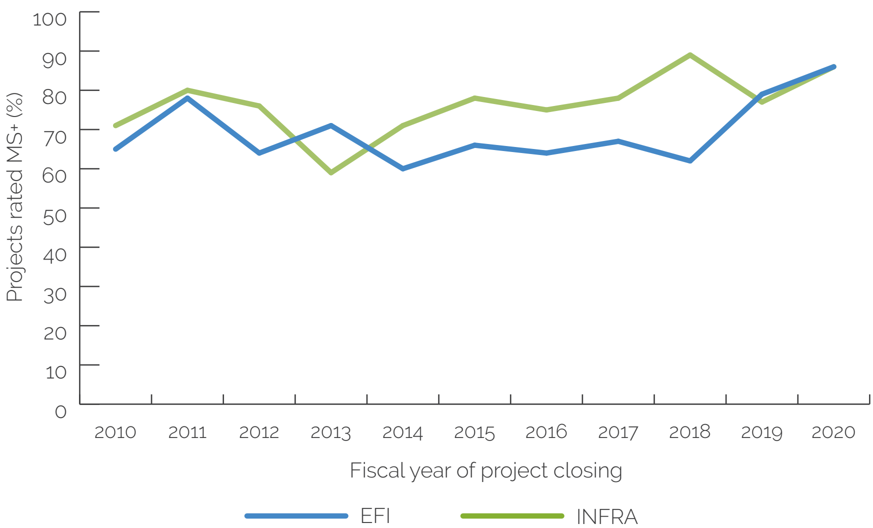



*Source:* Independent Evaluation Group.

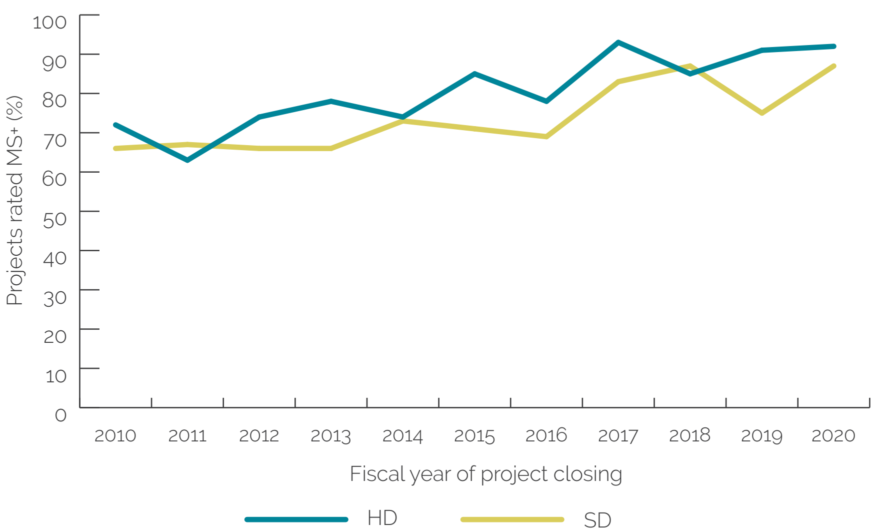
Ratings increased for projects in all Practice Groups, and the increase was especially steep for Sustainable Development projects. The percentage of Sustainable Development projects with project outcomes rated MS+ increased from 75 percent, the lowest among the four Practice Groups, in FY19, to 87 percent, the second highest, in FY20.[4] Sustainable Development projects also increased their share in the overall portfolio, from 38 to 48 percent of all projects. By contrast, the percentage of Human Development projects rated MS+ increased very little, from 91.1 to 91.7 percent, although Human Development still has the highest-rated projects of all Practice Groups (figure 2.3, panels a and b). This *RAP*'s decomposition analysis, presented in figure 2.4, shows that Sustainable Development's increase offsets the shrinking portfolio shares of the Infrastructure and Equitable Growth, Finance, and Institutions Practice Groups, which also had positive ratings increases.[5]

**Figure 2.3.** Projects Rated Moderately Satisfactory or Above for Select World Bank Project Categories

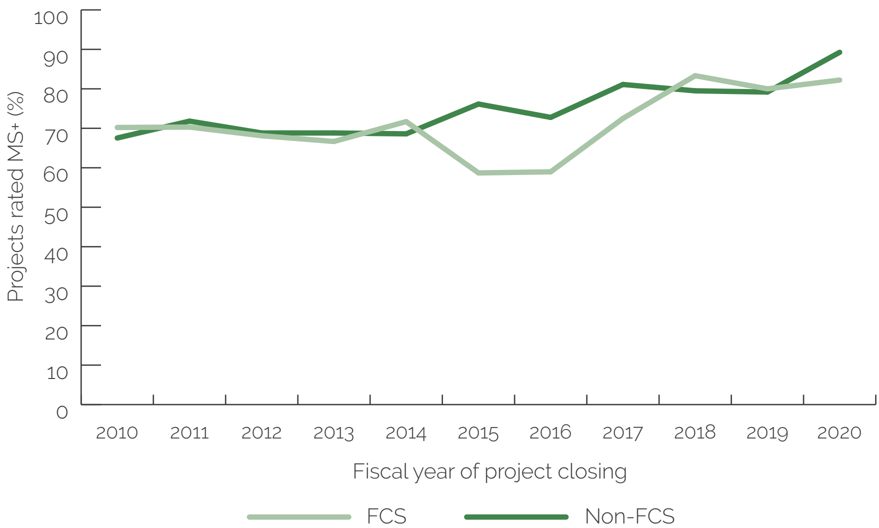
a. Practice Group: EFI and INFRA



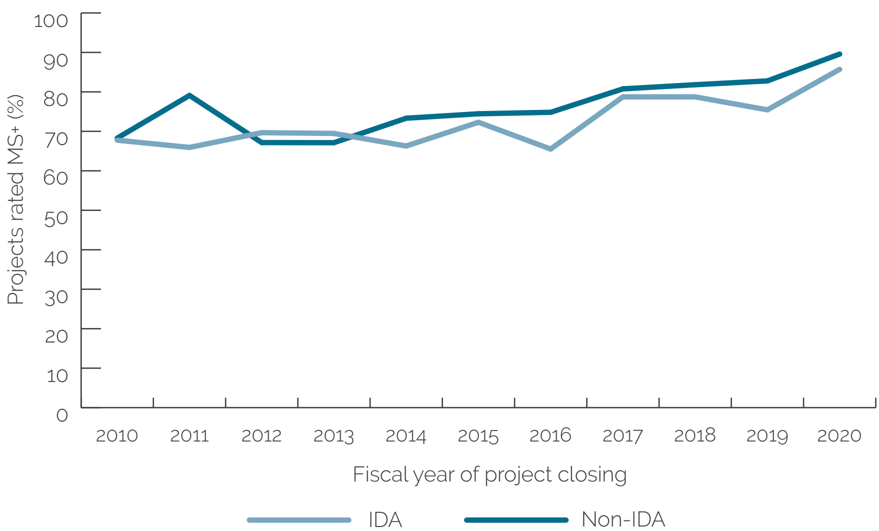

b. Practice Group: HD and SD



c. FCS status




d. IDA status




*Source:* Independent Evaluation Group.

*Note:* EFI = Equitable Growth, Finance, and Institutions; FCS = fragile and conflict-affected situation; HD = Human Development; IDA = International Development Association; INFRA = Infrastructure; MS+ = moderately satisfactory or above; SD = Sustainable Development.

All World Bank Regions experienced a large rating increase, but this was especially the case for Europe and Central Asia and Western and Central Africa. In Europe and Central Asia, projects rated MS+ jumped from 81 to 94 percent (+13 percentage points); Europe and Central Asia is now the Region with the highest percentage of projects rated MS+. In West Africa, the ratings increase was even bigger, jumping from 69 to 85 percent of projects rated MS+ (+16 percentage points). Moreover, both Regions expanded their relative share of projects. The limited number of projects in any single year's cohort, including FY20, prevented us from analyzing the ratings increase at the country level, but it is worth noting that the 34 projects in China, India, and Vietnam that closed in FY20 represent 18 percent of the World Bank's entire portfolio (by number of projects), and all 34 of these projects were rated MS+.

Ratings improved in FCS and IDA countries. Between FY19 and FY20, the percentage of projects with outcome ratings of MS+ increased from 80 to 82 percent in FCS countries and from 75 to 86 percent in IDA lending projects (figure 2.3, panels c and d). This is notable because both categories have an expanding share of World Bank projects. Between FY19 and FY20, the share of projects in FCS countries increased from 15 to 23 percent of the overall portfolio, whereas the share of IDA lending increased from 47 to 51 percent. At the same time, the project outcome ratings for non-FCS projects increased more (from 79 to 89 percent of MS+) than the ratings for FCS projects, which reopened the ratings gap—which had narrowed in FY19—between FCS and non-FCS countries. The increase in the percentage of projects rated S+ was steep for all project categories, but even more so for FCS and IDA lending countries. IDA project outcome ratings in FCS countries increased from 78 to 81 percent of MS+, and the share of these projects increased from 14 to 22 percent in the overall portfolio. These ratings increases are encouraging considering the World Bank continues to expand its engagement in both FCS and IDA countries.

Project outcome ratings also increased notably for the largest World Bank projects (those over $100 million). Past studies associate a project's size, by final project cost, with better ratings (Ralston 2014; World Bank 2016b).[6] The positive relationship between project size and outcome ratings intensified in FY20, when the percentage of large projects rated MS+ surpassed 95 percent.

At the same time, the outcome ratings for the smallest projects (below $10 million) decreased from 76 to 70 percent rated MS+.[7]

Disruptions from COVID-19 did not appear to explain the FY20 ratings jump. The *RAP*'s analysis dispels the concern that COVID-19's process disruptions might have biased the FY20 cohort toward more successful projects (see box 2.1).[8] Every year, projects with ICRs completed relatively quickly after the project closes tend to have higher ratings than projects whose ICRs are delayed, a pattern that the FY20 cohort displays nearly identically to previous cohorts.[9] However, in FY20 the ratings were higher than in FY19 throughout the year, including before the pandemic. It is also worth noting that none of the projects that closed in FY20 were prepared in response to COVID-19. Moreover, it is possible that work disruptions under COVID-19 and the need to channel emergency COVID-19 responses through existing projects led to an increase in the number of project extensions (from closing in FY20 to closing in future years). These extended projects will be accounted for in the fiscal year in which they actually close; they will be visible in, and possibly larger in, future ratings cohorts—and may affect those ratings—but they did not affect the FY20 ratings.

### Box 2.1. The Impact of the Coronavirus (COVID-19) Pandemic on Ratings

#### Potential Impact of Process Disruptions

Process disruptions related to the World Bank's response to the coronavirus (COVID-19) pandemic do not appear to have driven the recent ratings jump. During the COVID-19 pandemic, Implementation Completion and Results Reports (ICRs) arrived at intervals similar to those of previous fiscal years. World Bank operational teams experienced significant changes in working conditions after the COVID-19 pandemic began, but these changes did not appear to affect the normal inflow of ICRs to the Independent Evaluation Group (IEG). In most fiscal years, IEG sees a characteristic pattern in the arrival of ICRs, with large peaks near the end of December and June (quarters 2 and 4), and smaller peaks near the end of September and March (quarters 1 and 3). Since March 2020, however, the monthly pattern of ICRs arriving to IEG remained similar to that of previous years. Overall, there was a decrease in the number of ICRs, a trend that has persisted for several years. The current set of projects

*(continued)*

## Box 2.1. The Impact of the Coronavirus (COVID-19) Pandemic on Ratings (cont.)

validated by IEG in the fiscal year (FY)20 cohort is 194, with an additional 20 projects in IEG's validation pipeline and another 26 projects for which ICRs had not yet been received as of the cutoff date for this *Results and Performance of the World Bank Group*. The FY20 coverage (91 percent out of 214, or 81 percent out of 240) is in line with previous years. Refer to appendix A for additional methodological details.

Ratings in FY20 are higher than ratings in FY19, irrespective of how quickly ICRs and ICR Reviews (ICRRs) were completed. This *Results and Performance* report's analysis found that outcome ratings were negatively correlated with the length of time between the project's close and IEG's completion of the ICRR. This indicates that projects with higher ratings also completed their ICRs and ICRRs faster than projects with lower ratings. The analysis estimates a similar negative relationship for projects that closed in FY20 and projects that closed in previous years. However, the ratings for FY20 were higher than in FY19 throughout the self-evaluation and validation cycle compared with previous fiscal years.

### Potential Impact of Change in Evaluator Rating Standards

There is no evidence that outcome measurement standards were more lenient during FY20. The disconnect between ICR ratings and the ICRR ratings during COVID-19 appears similar to the disconnect of previous years. The disconnect is the difference between the outcome rating the World Bank assigns in the ICR self-evaluation and the outcome rating IEG assigns in the ICRR validation. The *Results and Performance* report's analysis of disconnect patterns shows fewer disconnects in FY20 (15 percent) than in FY19 (18 percent), but the difference is very small.

Ratings of ICR quality are roughly similar for recent and past ICRs. An internal IEG analysis of ICRRs completed recently indicates that ratings of ICR quality were similar to those reported in another IEG analysis from 2018. In both analyses, the outcome ratings assigned to projects with higher-quality ICRs (as measured by their ICR quality ratings) were less frequently downgraded (assigned a lower rating in the ICRR than in the ICR), and the pattern was similar in both periods.

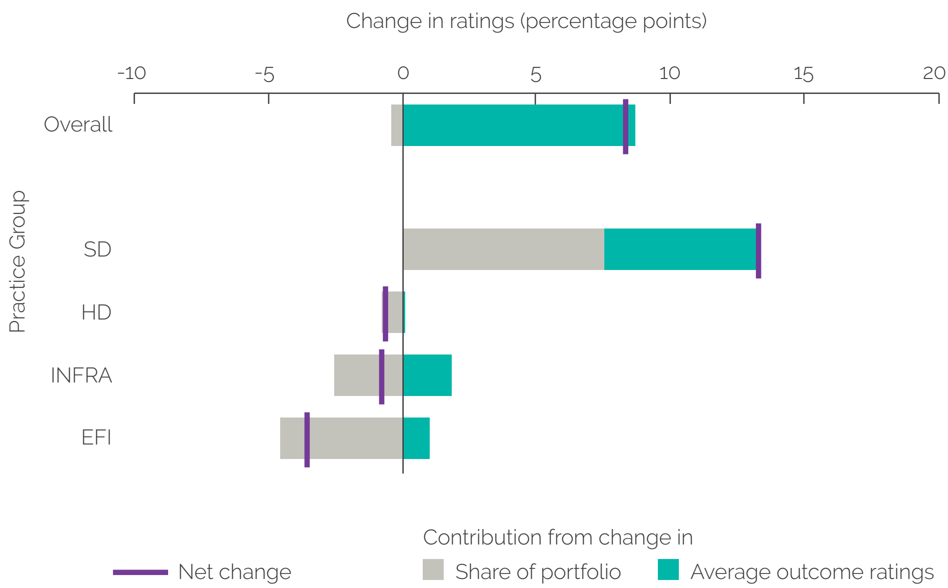
*Source:* Independent Evaluation Group.

The slower-than-usual processing time of development policy financing (DPF) for the FY20 cohort was unlikely to have affected the ratings increase. The World Bank's initiative to reform DPF ICRs, which began in 2017, led to a new DPF ICR template and a delay in processing both ICRs and ICRRs in FY20. Therefore, the number of FY20 DPFs whose ratings IEG validated is very small (only 9 of the 194 projects in the FY20 cohort). As a result, DPFs were unlikely to have contributed significantly to the overall increase in ratings, despite their strong ratings (all MS+). Moreover, there is no reason to expect that the validated ratings of processed DPFs will be significantly different from the ratings for DPFs that have not yet been validated.[10]

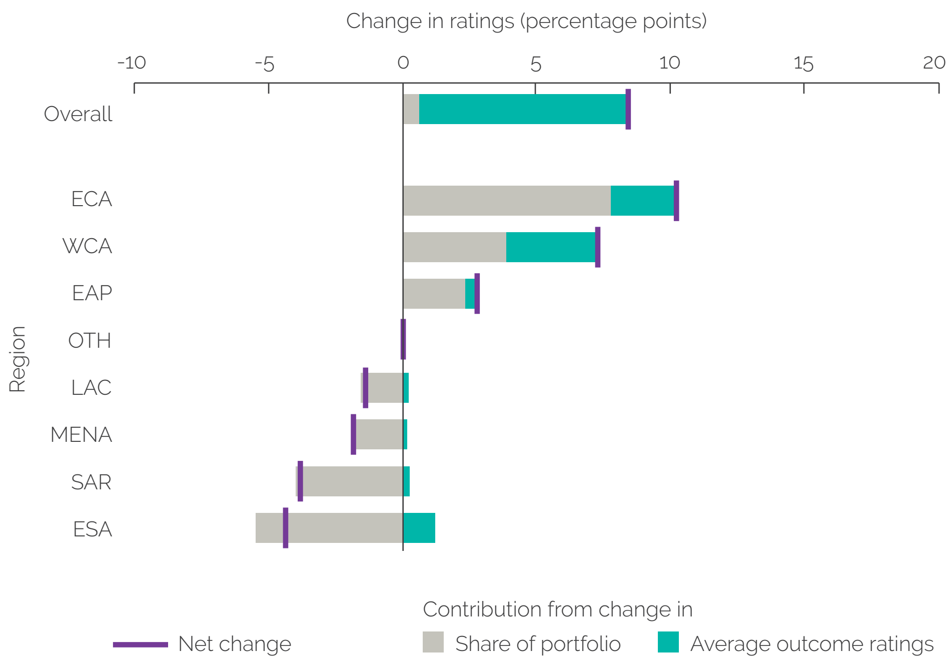
The FY20 increase in outcome ratings is notable because it occurred across almost all project categories. Ratings may increase because ratings for all project categories increase or because the share of already highly rated project categories in the overall portfolio increases. A decomposition analysis shows that the main contribution to the overall increase came not from portfolio changes but from ratings increases for virtually all categories of projects (figure 2.4).[11] For example, the large increase in Sustainable Development ratings combined with the simultaneously large increase in the proportion of Sustainable Development projects in the overall portfolio accounted for a disproportionately large contribution from this Practice Group to the overall change in ratings. However, this contribution was partially offset by the shrinking portfolio shares of Infrastructure and Equitable Growth, Finance, and Institutions, since both of these Practice Groups also had positive ratings changes. Therefore, despite some portfolio changes, it was the positive and often large ratings increases across nearly all project categories that drove the ratings jump. See appendix A for more details on the decomposition analysis.

## Figure 2.4. Select Contributors to the Increase in World Bank Project Outcome Ratings between Fiscal Years 2019 and 2020

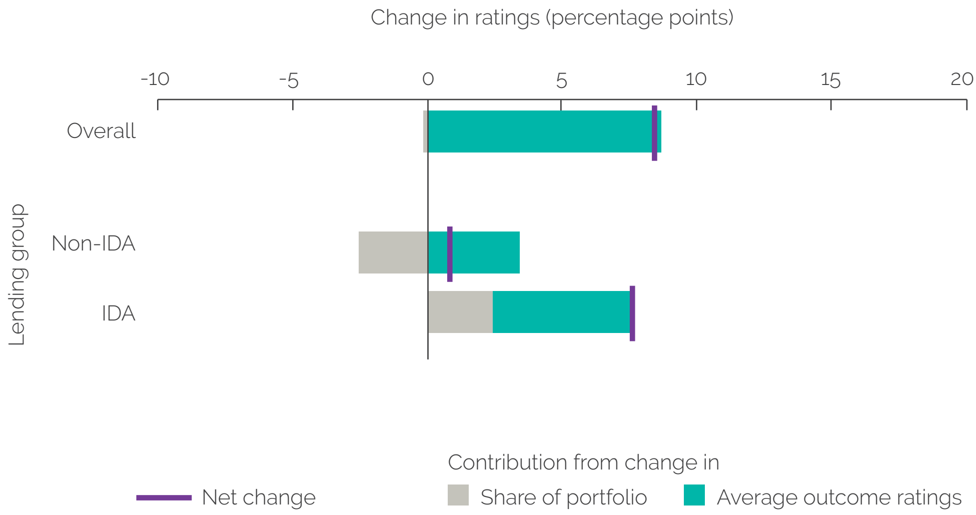
a. Decomposition by Practice Group



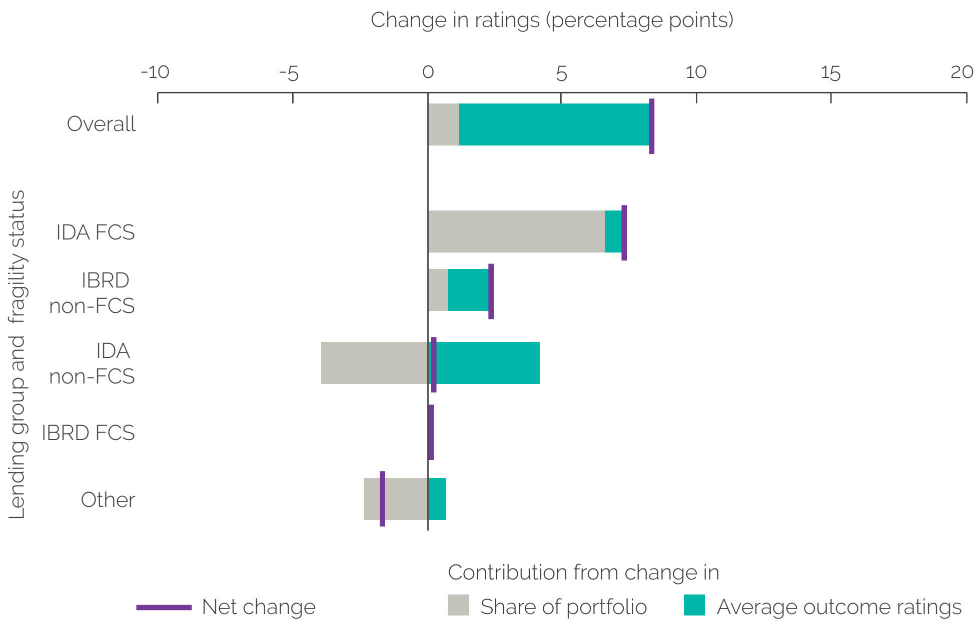

b. Decomposition by Region



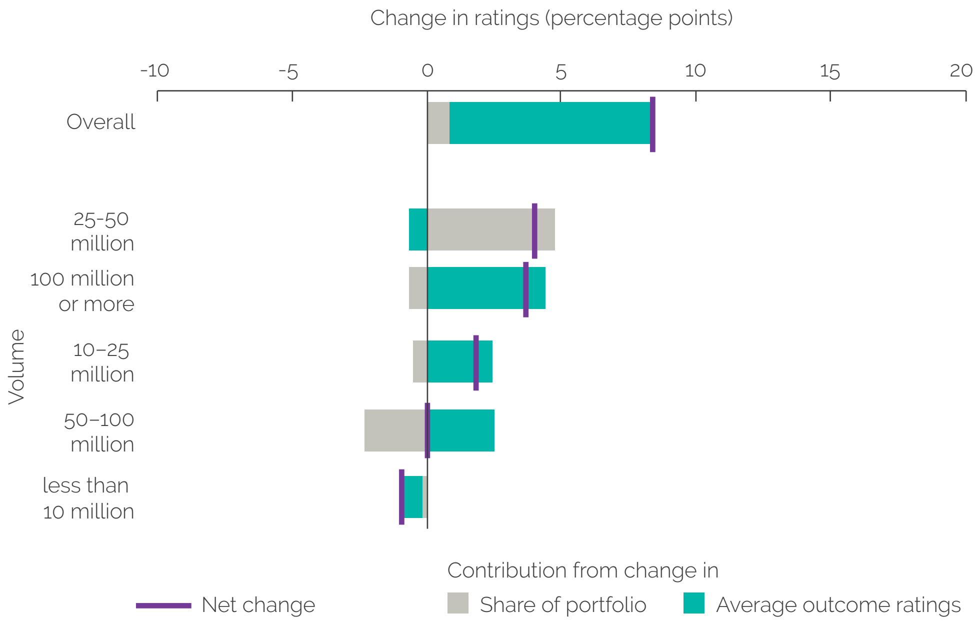
c. Decomposition by lending group




d. Decomposition by lending group and FCS status



e. Decomposition by volume




*Source:* Independent Evaluation Group.

*Note:* Net change refers to the net change in the overall average outcome rating between FY19 and FY20 (weighted by the share of projects in each category); contribution from change in share of portfolio refers to the contribution to the net change from changes in the share of portfolio of each category of projects across the two periods (the "between" effect); contribution from change in average outcome rating refers to the contribution to the net change from changes in average outcome ratings in each category across the two periods (the "within" effect). EAP = East Asia and Pacific; ECA = Europe and Central Asia; EFI = Equitable Growth, Finance, and Institutions; ESA = Eastern and Southern Africa; FCS = fragile and conflict-affected situation; FY = fiscal year; HD = Human Development; IBRD = International Bank for Reconstruction and Development; IDA = International Development Association; INFRA = Infrastructure; LAC = Latin America and the Caribbean; MENA = Middle East and North Africa; OTH = other; SAR = South Asia; SD = Sustainable Development; WCA = Western and Central Africa.

Bank performance ratings, which are positively correlated with the World Bank project outcome ratings, increased by 5 percentage points between FY19 and FY20, with 89 percent of projects rated MS+ in FY20. We cannot conclude that this increase drove the FY20 outcome ratings increase, but Bank performance ratings themselves are worth monitoring because they focus on elements within the World Bank's control.[12] Bank performance, defined in World Bank (2020b) as "the extent to which services provided by the World Bank ensured the operation's quality at entry and supported effective implementation through appropriate supervision (including ensuring adequate transition arrangements for regular operation of supported activities after closing)," is composed of two elements that also have their own ratings—a project's quality at entry and its quality of supervision.[13, 14] Many project categories that experienced large increases in project outcome ratings also experienced increases in Bank performance ratings—this was the case for Sustainable Development, Europe and Central Asia, IDA non-FCS projects, and especially large projects. However, for other project categories, such as projects in Western and Central Africa, project outcome ratings increased despite decreasing Bank performance. That said, it is likely that Bank performance affects the achievement of project outcomes, but since both ratings are assigned at the same time, causality cannot be determined.

Within Bank performance ratings, quality at entry ratings increased substantially between FY19 and FY20, possibly contributing to the outcome ratings jump. Past studies by IEG and others indicate that quality at entry ratings are strongly associated with outcome ratings (Chauvet, Collier, and Duponchel 2010; Raimondo 2016; Smets, Knack, and Molenaers 2013; World Bank 2015, 2016b). Considering this, quality at entry has been on an upward trend since 2014 and further increased from 76 percent MS+ in FY19 to 81 percent in FY20.[15] By contrast, the other half of Bank performance ratings—quality of supervision—has remained flat since FY18 after previous increases, albeit at a higher rate of 89 percent MS+. The *RAP 2018* found that the most important enablers of quality at entry were strong client relationships; well-timed analytical work as a foundation for project design; team composition and experience; and adequate internal policies, guidance, and systems (World Bank 2019c). The presence of key World Bank staff in the country also improved quality at entry, whereas the absence of staff hindered it. The *RAP 2018* identified challenges to quality at entry, including difficult operating environments (particularly in fragile situations), overly optimistic implementation schedules, weakly specified results frameworks, and projects not ready for implementation (for example, projects in which, at the time of approval, political economy issues were present, an implementation capacity assessment was not available, or questions related to a borrower's commitment or mandate for project execution were unresolved).

M&E quality ratings improved substantially between FY19 and FY20 and could partially explain the ratings jump. M&E quality generally rates low, but it has improved since FY10, when only 25 percent of projects were rated substantial or high. The percentage of projects with M&E quality ratings of substantial or high increased from 51 (in FY19) to 59 percent (in FY20). Project categories for which outcome ratings increased the most also showed increases in M&E quality ratings. There were large increases in M&E quality ratings in Sustainable Development, Water, and large projects—all of which had significantly higher ratings in FY20. That said, there were a few notable exceptions to the positive correlation between the project outcome rating and the M&E quality rating. For example, M&E quality decreased in all FCS countries, including IDA FCS countries, and in projects in the $25–50 million range, despite those project categories also having improved outcome ratings. M&E quality ratings also decreased in Human Development and the South Asia Region. A deeper analysis of M&E ratings, discussed in the next section, shows that the robust increase in M&E quality has the strongest positive correlation with the increase in project-level efficacy ratings, which—along with relevance and efficiency ratings—determine project outcome ratings.

Other country- and project-specific factors that IEG has previously identified continue to influence project outcomes, but these factors do not explain the large FY20 increase. Previous IEG regression analyses reveal that a country's capacity, a project's duration, a project's size, and the validation processing time all influence outcome ratings, and they continued to do so in FY20 (see World Bank 2018b, 2019c for a review of these factors). However, these factors' influence on FY20 outcome ratings was very similar to their influence on previous years' outcome ratings, so they do not explain the recent jump (appendix A, table A.1).[16]

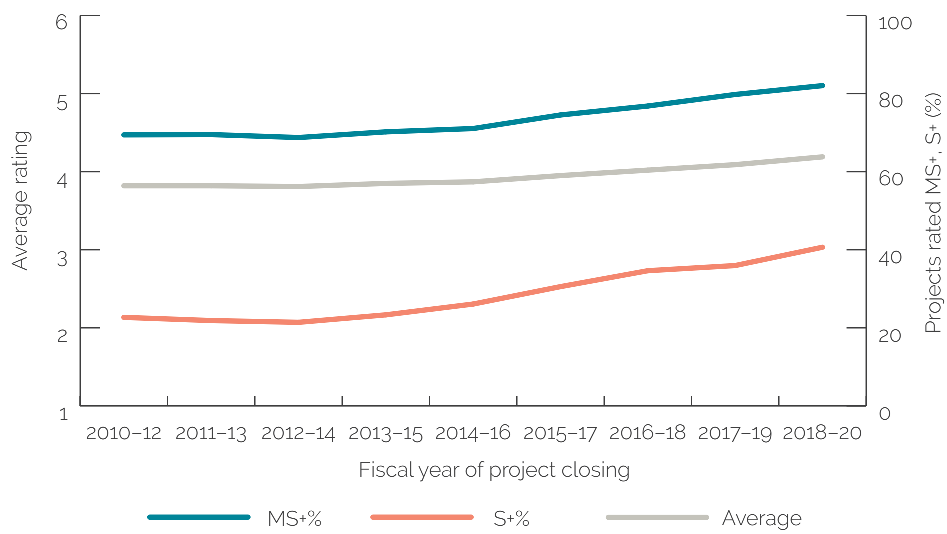
Other potential factors may help explain the ratings increase, but we could not measure these factors. These factors include teams improving measurement frameworks, preparing better theories of change, and becoming comfortable with the newer ICR methodology, among others. These improvements could have resulted from ICR reforms in 2017 or internal training and informal knowledge exchanges, which led to an increased focus by operational teams and development effectiveness units on building robust theories of change and paying more systematic attention to M&E quality. It is possible these factors came to fruition in FY20 and contributed to the ratings increase, but we could not measure whether this was the case.

## Longer-Term Ratings Improvements

The recent ratings increase continues a positive trend from previous years. This *RAP*'s analysis shows that longer-term ratings increases can occur with improved M&E and selective risk taking derived from adding new activities in successor projects. That said, this section also shows that the World Bank sometimes selects and approves indicators and targets that do not robustly measure development results.

Project outcome ratings have been increasing over the past seven to eight years. A three-year rolling average shows that starting from FY13, ratings have steadily increased, both for average outcome ratings and as a percentage of projects with outcomes rated MS+. The percentage of projects rated S+ has also been markedly increasing (figure 2.5).

Figure 2.5. World Bank Project Outcome Ratings, Three-Year Rolling Average, by Number of Projects




*Source:* Independent Evaluation Group.

*Note:* MS+ = moderately satisfactory or above; S+ = satisfactory or above.

## Ratings and Project Novelty

The longer-term upward trend in ratings does not appear to be due to less risk being taken. We looked at aversion to risk in terms of the tendency to repeat project designs in successive projects versus the willingness to expand or introduce new elements in successor project designs. We focused on two Global Practices—Education and Transport—with larger portfolios and large shares of highly rated projects (see a methodological summary in box 2.2). Contrary to expectations, successor projects that introduced novelty, or new elements, performed as well as or better than projects that closely replicated their predecessors. This indicates that the World Bank has been able to take informed risk and selectively introduce new elements relevant to context without suffering lower project outcome ratings. That being said, the analysis did not directly look at projects' contextual elements or task teams' risk-mitigation measures; moreover, the evidence from this analysis is from only two sectors and not entirely representative of how World Bank projects, as a whole, build on and learn from past projects.

### Box 2.2. Methodology of the Novelty Analysis

We carried out a detailed analysis of how consecutive Transport and Education projects changed in individual countries. We selected a sample of projects that closed in fiscal year (FY)12–14 (early period) and FY17–20 (later period).[a] The sample was representative of the Transport and Education Global Practices' project portfolios that closed in each period, with a 90 percent confidence level and a 10 percent margin of error. The team matched each sampled project to a predecessor project, defined—for the purposes of this exercise—as a project in the same Global Practice and the same country that closed five years or less before the sampled successor project started; covered at least one common subsector; and had the same implementation agency as the predecessor project. In total, the sample included 75 Transport projects (of which 49 had a predecessor) and 79 Education projects (of which 55 had a predecessor). We analyzed these projects' activities using Implementation Completion and Results Reports, Project Appraisal Documents, and Implementation Completion and Results Report Reviews as sources of information.

We compared predecessor and successor projects by looking at the projects' Independent Evaluation Group ratings, the number of new activities, and their level of difficulty as determined by Independent Evaluation Group sector experts.[b] We rated the novelty of each successor project as limited, moderate, or high, according to the following criteria:

(continued)

## Box 2.2. Methodology of the Novelty Analysis (cont.)

- Successor projects of limited novelty had up to two new activities that the predecessor project did not have, at least one of which was assessed to be of limited difficulty, and the project's scope remained roughly the same or was scaled down.
- Successor projects of moderate novelty had between two and four new activities that the predecessor project did not have, at least half of which were assessed to be of moderate difficulty, or had more than four new activities that were assessed to be of limited or moderate difficulty.
- Successor projects of high novelty had more than four new activities that the predecessor project did not have, at least two of which were assessed to be of high difficulty.

The team also examined first-time projects, which did not have predecessors. This group of projects included 26 first-time projects for Transport and 24 first-time projects for Education. Appendix A provides more details on this analysis.

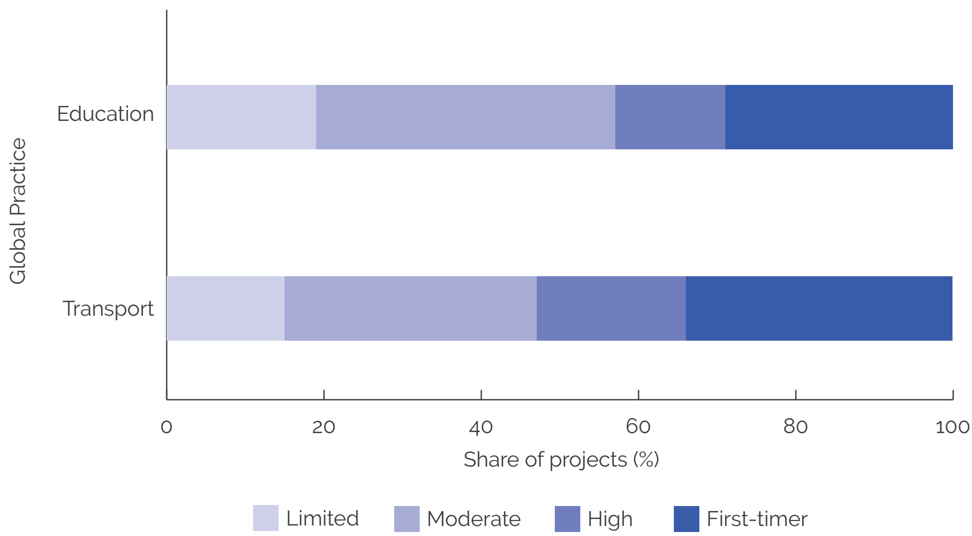
*Source:* Independent Evaluation Group.

*Note:* a. The samples for the two periods were pooled because no differences in rating patterns were detected across the periods.
b. In Education, some activities with limited difficulty included school construction, teacher training, or textbook policies and distribution. Activities of moderate difficulty included curriculum reviews, performance-based agreements and revision of work programs, or teacher certification systems. Activities of high difficulty included sector reforms, governance systems, or legal frameworks. In Transport, activities of limited difficulty included infrastructure rehabilitation and maintenance, feasibility studies, or staff training. Moderately difficult activities included highway, railway, or airport construction; the introduction of performance-based contracts; or the revision of public-private partnership frameworks. Highly difficult activities included establishing sector agencies, setting up a road fund, or designing and implementing sector reforms.

The World Bank tends to build on previous projects rather than introduce entirely new projects, in both the Transport and Education Global Practices (GPs). Approximately one-third of projects in both GPs were first-time projects that did not have a predecessor project (29 percent in Education and 35 percent in Transport), whereas the remaining two-thirds of projects were linked to predecessor projects (figure 2.6).

Most successor projects introduced new elements that were not included in their predecessor projects; few projects repeated the predecessor's project design. Approximately half of the sampled successor projects in both GPs were moderately novel, whereas 20 percent of Education successor projects and 29 percent of Transport successor projects were highly novel. This shows that Education projects tend to introduce fewer novel elements than Transport projects, but overall the pattern of novelty is similar in both GPs. Likewise, only approximately one-quarter of successor projects in both GPs have limited novelty, or largely repeat project elements from their predecessor projects (figure 2.6).

Figure 2.6. Distribution of Education and Transport Projects, by Novelty Level



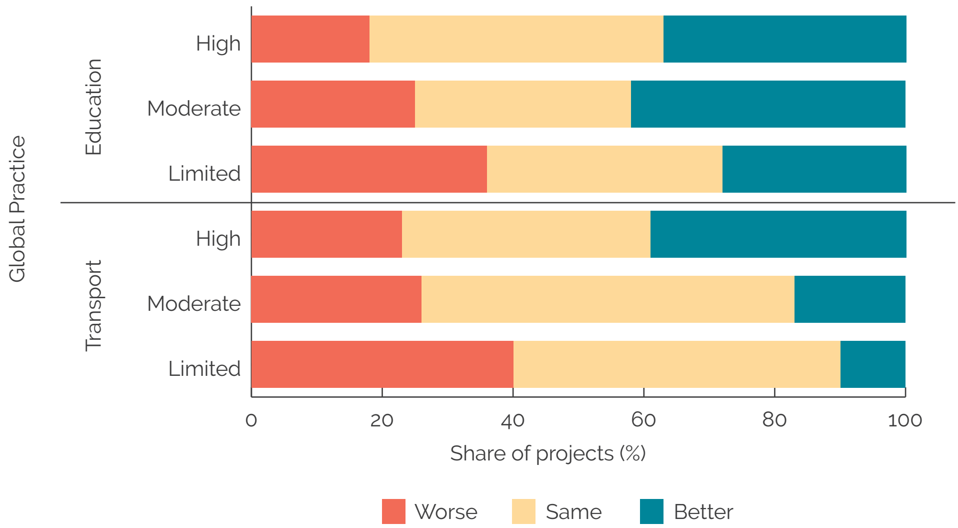

*Source:* Independent Evaluation Group.

The new elements in project designs consisted of more complex and challenging activities or a broader geographic or thematic scope than the predecessor project. In an example of greater complexity, the Ghana Education For All Fast Track Initiative (P116441) provided learning materials, salary incentives, and teacher training to attract teachers to deprived districts. The follow-on Ghana Partnership for Education project (P129381) carried out similar activities but added elements of institutional strengthening (such as annual work programs with district education offices and school improvement grants), which increased the project's level of difficulty because the success of these elements often depends on external factors and has long time horizons. In an example of greater geographic scope, the predecessor Buenos Aires Urban Transport Project (P039584) financed public transport interventions and technical assistance in the Buenos Aires metropolitan area. The successor Argentina Urban Transport in Metropolitan Areas project (P095485) then extended the project's scope to medium-size cities in Argentina. In an example of greater thematic scope, the Road Sector Development Project (P050623) in Ghana was essentially an interurban road rehabilitation and improvement project. Meanwhile, the follow-on Transport Sector Project (P102000) ventured into new subsectors, such as air, urban, and waterborne transport, and had several additional goals in new thematic areas, including improved sector planning, upgraded urban roads and urban transport infrastructure, and strengthened institutional capacity for transport agencies.

The projects that introduced new elements, or novelty in project design, achieved higher increases in ratings than projects that simply repeated their predecessors. In Education, 82 percent of successor projects that introduced high novelty performed better than or the same as their predecessors, whereas only 64 percent of projects that introduced limited novelty performed better or the same. Something very similar happened in Transport (figure 2.7). Among projects that introduced greater novelty, most of their predecessors were rated MS+. In fact, there was only one example from each GP where a task team added a high level of novelty to a project that followed a project rated unsatisfactory. There is no evidence that task teams use novelty to course correct unsatisfactory projects. Rather, task teams phase novelty into projects that already perform well and can more easily sustain expanded, or more difficult, activities while maintaining or even increasing ratings. This suggests that World Bank task teams introduce novelty to more promising projects, hence taking risks when conditions indicate a higher likelihood of success.

First-time projects were simpler—that is, had fewer elements—than successor projects in the Education and Transport GPs. In Education, first-time projects, on average, covered fewer subsectors, had fewer safeguard implications, and had fewer commitments compared with successor projects.[17] In Transport, first-time projects had fewer project sites, subsectors, and commitments, but had more safeguard implications. By designing low-complexity first-time projects, task teams decrease the risk and uncertainty of those projects.

**Figure 2.7.** Changes in World Bank Project Outcome Ratings from Predecessor to Successor Projects, by Global Practice and Novelty Level




*Source:* Independent Evaluation Group.

*Note:* Red indicates successor projects with lower ratings than their predecessor project; yellow, successor projects with the same ratings as their predecessor project; and teal, successor projects with higher ratings than their predecessor project.

## Ratings and Types of Outcomes

A project's M&E quality plays a larger role in explaining ratings than the type of outcomes a project aims to achieve. We analyzed whether ratings vary when project outcomes, defined as the intended changes pursued by a specific project objective, vary (box 2.3).[18] Our analysis indicates that there is a weak relationship between a project's outcome types and its objective-level efficacy ratings (see appendix B for details on the methodology and the outcome type analysis). Only 4 outcome types out of the 16 identified—expanded access to services, increased human capital, improved enterprise and sector performance, and enhanced equity and inclusion—were associated with higher objective-level efficacy ratings than the remaining outcome types. However, this was largely due to the fact that the projects pursuing these outcomes had stronger M&E. That is, the M&E quality of a project plays a bigger role in explaining ratings than do outcome types. To some extent, this is likely due to the rating methodology: The project's achievement is assessed against the project's specific targets and objectives rather than against standardized objectives and targets associated with standardized outcome types.

### Box 2.3. Methodology of Outcome Type Analysis

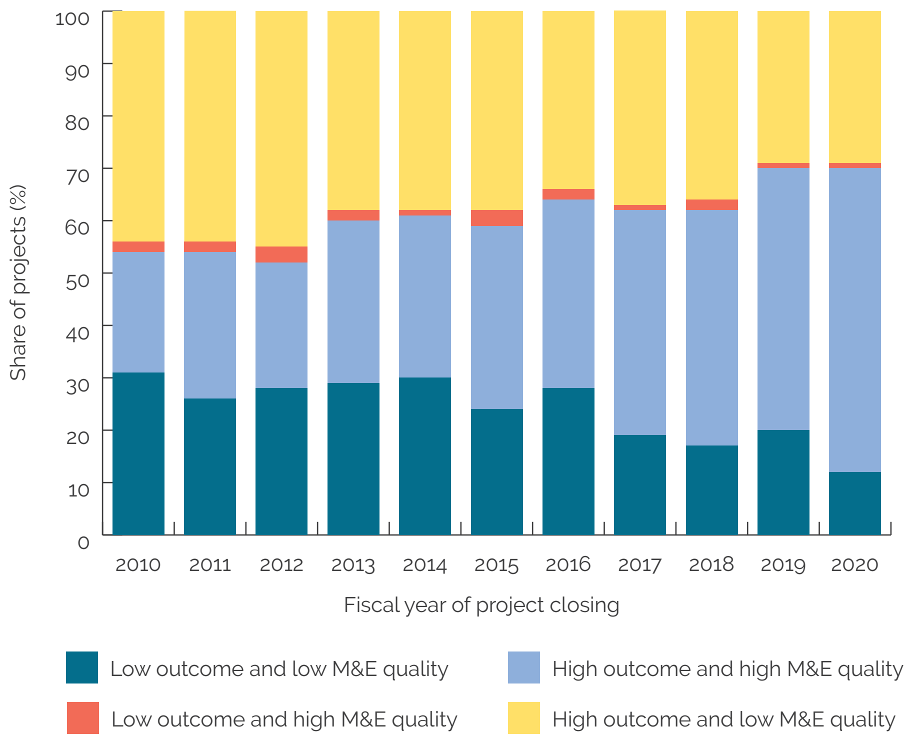
*Outcome types* capture the type of change envisioned by project objectives. We defined an outcome typology, which includes 16 outcome types, derived from typical project theories of change and select corporate objectives. Examples of outcome types include expanded access to services, enhanced institutional capacity, improved service quality, increased human capital, and 12 others. We analyzed outcome types for a representative sample of World Bank projects for two different time periods: projects that closed in fiscal year (FY)14–16 and projects that closed between FY17 and the second quarter of FY20. The two samples were representative at the Practice Group level with a 90 percent confidence level and a 10 percent margin of error. We used each individual project objective identified during validation as the unit of analysis. This allowed the team to analyze variations in objective-level efficacy ratings as outcome types vary. Appendixes A and B contain additional details.

*Source:* Independent Evaluation Group.

## Ratings and Monitoring and Evaluation

Better M&E helps explain higher project outcome ratings. The percentage of projects with M&E quality rated substantial or high has increased considerably, from 25 percent in FY10 to 59 percent in FY20. The largest increases have occurred since FY16, when project outcome ratings also increased the most. Also, the share of coherently rated projects—that is, those with high M&E quality ratings and high project outcome ratings, or low M&E quality ratings and low project outcome ratings—has also increased since 2010 (figure 2.8). The percentage of coherently rated projects increased from 54 percent in 2010 to 70 percent in both FY19 and FY20. Although there is no proof of causation, studies have suggested that M&E quality may have a genuine impact on project outcome ratings. For example, Raimondo (2016) calculates that projects with high M&E quality ratings perform between 0.13 and 0.40 points better (in terms of average project outcome ratings) than projects with low M&E quality. One potential explanation is that better M&E systems allow better adaptive management, which is necessary to ensuring project success.

**Figure 2.8.** World Bank Project Outcome Ratings and Monitoring and Evaluation Quality




*Source:* Independent Evaluation Group.

*Note:* "High" M&E quality corresponds to substantial or high ratings. "Low" M&E quality corresponds to modest or negligible ratings. M&E = monitoring and evaluation.

## Ratings and Project Indicators and Targets

High project outcome ratings do not necessarily mean that projects have higher quality indicators or more ambitious targets. First, in FY20, 29 percent of projects with outcomes rated MS+ had only modest or negligible M&E quality ratings (figure 2.8). Second, an in-depth analysis of the type and quality of indicators and targets showed that even when indicators and targets are not fully adequate for measuring results, projects can still achieve good ratings (see box 2.4). This latter finding in particular suggests that project teams and operational management do not systematically scrutinize the selection of project indicators and targets, which as a result differ widely in their adequacy to measure project achievements directly and objectively. At the same time, projects that meet their defined targets are likely to achieve higher ratings regardless of what those targets represent.

## Box 2.4. Methodology of the Indicator and Target Analyses

To explore the relationship between outcome ratings and monitoring and evaluation, we analyzed the type and quality of indicators in projects with institutional strengthening objectives and all the targets of project development objective indicators of a sample of projects with only high, or only low, objective-level efficacy ratings.

**Indicator analysis.** The *Results and Performance of the World Bank Group* team analyzed the project objectives characterized as having "institutional strengthening" outcome types (appendix B) using projects sampled for the novelty and outcome type analysis (appendix A). We selected projects with institutional strengthening objectives because this is one of the most prevalent outcome types, is present in all sectors, and is critical for achieving development results (including the sustainability of development results). The analysis included objectives that were solely institutional strengthening and objectives that were institutional strengthening along with other outcome types. We analyzed 707 objectives in 268 projects. We identified, classified, and linked the institutional strengthening indicators to their corresponding objective-level efficacy ratings based on information from Implementation Completion and Results Report (ICR) Reviews.

**Target analysis.** We analyzed the achievement of project development objective indicator targets using projects sampled for the novelty and outcome type analysis (appendix A). Among those projects, we selected projects in both extremes of the objective-level efficacy distribution—that is, projects in which either (i) all the objective-level efficacy ratings were negligible or modest (76 projects), or (ii) either the objective-level efficacy ratings were all high, or some were high and others substantial (39 projects). The 115 projects encompassed 647 project development objective indicators and targets. Next, we excluded indicators whose targets were revised, so we could maintain an unambiguous connection with objective-level efficacy ratings. This resulted in a sample of 340 indicators across 79 projects. We classified each target as exceeded, fully achieved, partially achieved, not achieved, or no evidence provided. The unit of analysis was the indicator (along with its associated target). We used ICR annexes instead of ICR Reviews as the primary source of information because ICR annexes, unlike ICR Reviews, list project development objective indicators and targets systematically. Selecting projects with all individual objectives rated negligible or modest or all rated

*(continued)*

## Box 2.4. Methodology of the Indicator and Target Analyses (cont.)

high (or some high and some substantial) helped to unequivocally associate each efficacy rating with a single achievement category (exceeded, fully achieved, and so on), but this implies that the analysis is not generalizable to all projects.

Appendix A provides more details on both analyses.

*Source:* Independent Evaluation Group.

Not all individual project objectives have indicators to measure them, and many rely on weak evidence. We analyzed in-depth projects with institutional strengthening–related objectives and found that 7 percent had no defined indicators to measure them.[19] Moreover, 53 percent have indicators that measure only outputs or rely on anecdotal evidence (table 2.1). Output indicators measure whether teams completed actions toward achieving an outcome rather than whether the outcome has been achieved. Efficacy ratings based on output indicators can therefore be rated substantial or high even when the project did not achieve its institutional strengthening outcome. Something similar happens when projects measure objectives with anecdotal evidence, which is considered a weak form of evidence when it relies on personal observations collected in a nonsystematic manner. For example, projects used the opinion of individual stakeholders, who were not purposefully selected, to determine whether a project achieved its institutional strengthening objectives rather than using a more systematic, independent, or objective measurement.[20] These findings suggest room for improvement in ensuring adequate attention to results measurement and to the quality and appropriateness of indicators (table 2.1).

Moreover, projects often measure institutional strengthening objectives only indirectly. For 28 percent of institutional strengthening–related objectives, the project did not directly measure institutional strengthening (table 2.1). Rather, these projects measured the potential consequences of having stronger institutions, in terms of, for example, reduced travel times, reduced emissions, decreased rates of maternal deaths, improved health behaviors, improved education participation, and so on. It is possible that institutional strengthening activities did contribute to these outcomes, but this contribution was not measured. This is an example of low construct validity (that is, indicators poorly measuring what they are supposed to measure). In such cases, it would be possible for institutional strengthening objectives to receive high efficacy ratings without achieving the actual outcome, or conversely for projects that achieve their institutional strengthening objectives not to receive high ratings.

**Table 2.1.** Quality of Indicators Measuring Institutional Strengthening Objectives

| Approach to Indicators | Share of Objectives Pursuing Institutional Strengthening (%) | Average Objective Efficacy Ratings |
|---|---|---|
| No indicator defined | 7 | 2.0 |
| Weak indicator definition (outputs or anecdotal evidence) | 53 | 2.6 |
| Indicator defined in terms of indirect measurement only | 28 | 2.8 |
| Indicator defined in terms of relatively more direct or "plausible" measures of institutional strengthening | 30 | 2.7 |

*Source:* Independent Evaluation Group.

*Note:* The analysis is based on 707 objectives in 268 projects. The categories above are not mutually exclusive. The average efficacy ratings have been calculated based on the following scale: high = 4, substantial = 3, modest = 2, negligible = 1.

That said, many projects adequately measure institutional strengthening objectives. Approximately one-third of objectives used a more direct or "plausible" measurement approach (table 2.1). In projects with institutional strengthening objectives, a direct approach measures the performance of the institutions that the project strengthened. For example, direct indicators included increased ministry revenues or expenditures; decreased time for the institution to process licenses or disseminate annual statistics; increased on-time court case settlements; more project approvals by the relevant ministry; more reports published; and fewer inclusion errors from an agency's targeting mechanisms. Meanwhile, a plausible approach measures results that were plausibly attributable to institutional strengthening activities. In some cases, this approach measures demand-side factors, such as beneficiary satisfaction with an institution or training program, or supply-side factors by assessing a ministry's capacity or training participants' skill levels. Box 2.5 provides examples of robust indicators that projects have used to measure institutional strengthening.

### Box 2.5. Examples of Project Objectives with Robust Measurement of Institutional Strengthening

#### Second Eastern Indonesia Region Transport Project (P074290)

Objective: Decentralize the planning and management responsibilities for public works on Indonesia's provincial and *kabupaten* (regency) roads from the central to the provincial and *kabupaten* governments.

Measurement: The project measured this objective with supplemental performance evaluation data on timeliness, documentation, and successful completion of the intended civil works programs in each territory. The task team created an index of performance in which a score of 70 percent was the threshold for acceptable performance.

#### Punjab Municipal Services Improvement Project (P083929)

Objective: Improve the delivery and effectiveness of urban services in Punjab's participating municipalities in Pakistan.

Measurement: The project measured this objective using two institutional development assessments by the Punjab Municipal Development Fund Company (one of the two implementing agencies), with assistance from the World Bank's supervision team. The first assessment was completed just before the Mid-Term Review in January 2010, and the second was completed in early 2013, approximately six months before closure.

#### Maputo Municipal Development Program II (P115217)

Objective: Improve the sustainability of municipal services in Mozambique's Maputo Municipality.

Measurement: The project measured this objective using citizen report cards that recorded residents' overall perceptions of city services. The mean scores were reported to assess changes in resident perceptions of services.

*Source:* Independent Evaluation Group.

Objectives that lacked indicators had lower efficacy ratings, but projects with higher quality indicators did not always have higher efficacy ratings. Objectives with no indicators (no evidence) to measure institutional strengthening had an average efficacy rating of 2, measured on a scale from 1 to 4. Objectives that measured institutional strengthening more directly had an average efficacy rating of 2.7, whereas objectives that measured institutional strengthening indirectly or through weak evidence had ratings of 2.8 and 2.6 (table 2.1). That is, projects with more rigorous indicators do not have higher efficacy ratings than projects with less rigorous indicators.[21]

When projects use weak evidence to measure project objectives and still obtain relatively high ratings, one questions the reliability of efficacy ratings. The use of output indicators or unreliably collected anecdotal evidence to measure outcomes leads to validity issues and possibly to artificially high (or low) efficacy ratings. This is because weak evidence creates an arbitrary space for deciding whether or not projects achieved results.

Meeting targets is strongly associated, by design, with higher efficacy ratings. Among projects with highly rated objectives, the majority of targets were achieved or surpassed (table 2.2). Efficacy ratings decrease as the percentage of targets that a project achieves diminishes. That said, a notable share of projects with low (modest or negligible) efficacy ratings also achieved or surpassed their targets (table 2.2). This can happen in several scenarios: (i) When the project team writes the ICR, or when IEG completes the ICRR, they determine that one or more indicators underlying the targets were actually not adequate to measure objectives (for example, when project development objective indicators measure only activities, outputs, or beneficiaries rather than measuring the intended outcomes defined in the project development objective); (ii) there is evidence that the targets were intentionally overly conservative; or (iii) IEG's ICRR finds less or weaker evidence of achievement than the ICR does for the same indicator and targets (a circumstance that may be more likely when qualitative indicators are used).[22]

The *RAP*'s closer analysis of targets uncovered a number of shortcomings. First, a relatively high percentage of (quantitative) targets that were achieved were *exactly* achieved, an implausible result that calls into question the reliability of reported data. Twenty-seven percent of the original targets and 33 percent of the revised targets classified as "fully achieved" achieved their quantitative target exactly. Second, many targets are set in absolute terms, providing a measure of the project's size or reach, rather than in relative (percentage) terms, providing a measure of how well the project fills gaps or meets development needs. Third, many project development objective indicator targets had a zero baseline, often meaning that these indicators only measure outputs or lack data to build a baseline.[23]

**Table 2.2.** Indicators in Each Achievement Category of Project Development Objective Targets for Three Project Groupings (percent)

| Achievement of Indicator Targets | Grouping of Projects | | |
|---|---|---|---|
| | All Objectives Had High Efficacy Ratings | Objectives Had a Mix of High and Substantial Efficacy Ratings | All Objectives Had Negligible or Modest Efficacy Ratings |
| Exceeded | 54 | 42 | 26 |
| Fully achieved | 43 | 28 | 21 |
| Partially achieved | 3 | 15 | 21 |
| Not achieved | 0 | 6 | 18 |
| No evidence | 0 | 3 | 10 |
| Other | 0 | 6 | 4 |
| Total | 100 | 100 | 100 |
| Number of indicators | (*N* = 35) | (*N* = 67) | (*N* = 238) |

*Source:* Independent Evaluation Group.

*Note:* The analysis is based on projects in the two extremes of the objective-level efficacy ratings distribution selected from those sampled for the outcome type analysis; that is, it takes a representative sample of projects that closed in fiscal year (FY)12–14 and from FY17 to the second quarter of FY20 (see appendix A). From that sample, this analysis excludes indicators whose targets were revised to maintain an unambiguous connection between indicators-plus-targets and objective-level efficacy ratings. The resulting analysis set included 340 indicators. "Other" includes indicators that have no baselines or no targets, or were dropped.

Neither IEG nor World Bank management often scrutinize the appropriateness of project targets. This is surprising considering how fundamental achieving targets is to high efficacy ratings. IEG's objective-based methodology does not require IEG reviewers to scrutinize the quality of targets; therefore, ICRs rarely discuss or justify how targets are set or revised. And yet, if ratings strongly depend on targets being achieved, a legitimate question is whether these targets represent a sufficient improvement over baselines and are not too conservative. A better process for target setting that is more transparent could help IEG and the World Bank better assess projects' actual achievements.

[1] Ratings for World Bank projects are aggregated according to the closing fiscal year of the project. Ratings become available after three milestones occur: (i) the project closes; (ii) the project team completes theImplementation Completion and Results Report (ICR; usually within 6 months of project closing) and the Regional Director approves it, which results in the ICR arriving at the Independent Evaluation Group (IEG) for validation; and (iii) IEG completes the ICR Review (ICRR), thereby validating the ICR and producing final ratings. The cutoff date for trend data reported in this *Results and Performance of the World Bank Group* (*RAP*) is August 10, 2021. Thus, "FY20 projects" indicates World Bank projects that closed in fiscal year (FY)20, and completed their ICR and had their ICR validated by IEG in an ICRR by August 10, 2021. A total of 194 projects met this requirement and were therefore accounted for; by August 10, 2021, another 20 had not been validated yet and 26 had not completed an ICR. The number analyzed therefore represents 91 percent of all projects (or 194/214, if only those projects waiting for an ICRR are considered) or 81 percent (or 194/240, if projects missing an ICR are also included in the denominator). These are very high percentages compared to previous years.

[2] For a review of factors associated with World Bank project performance, see World Bank 2018b.

[3] There was an increase in project outcomes rated moderately satisfactory or above from 71 to 80 percent from FY16 to FY17, which was driven by a decrease in the percentage of projects with unsatisfactory ratings.

[4] This follows, however, a steep drop from FY18 to FY19.

[5] Water (+16.8 percentage points increase in projects with outcome rated moderately satisfactory or above) and Urban, Disaster Risk Management, Resilience, and Land (+20.5 percentage points) were the two Global Practices (GPs) within Sustainable Development that had the largest increase in ratings. Both GPs also increased their relative portfolio shares. However, it should be stressed that the number of projects within each GP for individual FYs is not large enough to meaningfully analyze ratings at the GP level, especially for the smaller GPs.

[6] Ralston (2014) found that large projects tended to have higher outcome ratings. IEG's *RAP 2015*, which analyzed projects based on the project's initial size compared with its final size, found that the change in project size is significantly correlated with project outcome ratings (World Bank 2016b).

[7] IEG's *RAP 2017* (World Bank 2018b) found that the distribution of World Bank financing volumes across projects is uneven, with a few very large projects accounting for a large share

of this volume. Among projects that closed during FY11–16, very large projects ($500 million or above) accounted for 4 percent of projects and 35 percent of volume, and projects under $100 million accounted for 74 percent of projects and 21 percent of volume. Projects between $100 million and $500 million accounted for 22 percent of projects and 44 percent of volume.

[8] For details on the number of projects still unaccounted for in the FY20 cohort, see endnote 1.

[9] This dispels the hypothesis that the increase in FY20 ratings was due to a disproportionate share of highly rated projects completing their ICR and ICRR cycle quicker than in previous years because of work disruptions from the coronavirus pandemic.

[10] Development policy financing  had historically similar or higher ratings than investment policy financing, except in FY17 and FY18, when the percentage of development policy financing rated MS+ was noticeably lower than for investment policy financing. In any case, the small amount of development policy financing in FY20 cannot impact the overall project outcome ratings meaningfully.

[11] The figure presents the results of a decomposition of the increase in project outcome ratings between FY19 and FY20, separately calculated for each category of projects. There is an online dashboard for this year's *RAP* that includes more project categories and shows results for other types of ratings besides project outcome ratings, such as monitoring and evaluation quality, World Bank quality at entry, and World Bank quality of supervision. Appendix A shows the formula to decompose the ratings. Because of a small *N* value, decomposition analysis is not meaningful at the GP level.

[12] Ultimately, outcome ratings are informed by project achievements that depend not only on the World Bank's efforts but also on external factors such as borrower performance, a country's economy, the government's political commitment to reform, and the general context and enabling environment.

[13] Quality at entry refers to how well the World Bank identified, prepared, and appraised a project so that it was most likely to achieve the project's planned development outcomes and was consistent with the World Bank's fiduciary role. Criteria for the rating of quality at entry include the following: the project's strategic relevance and approach; technical, financial, and economic aspects (for investment project financing); poverty, gender, and social development aspects; environmental aspects (including provisions for safeguard policy compliance); fiduciary aspects; policy and institutional aspects; implementation arrangements; monitoring and evaluation arrangements; risk assessments; and World Bank inputs and processes.

[14] Quality of supervision refers to the extent to which the World Bank proactively identified and resolved threats to the achievement of relevant development outcomes. Criteria for the rating of quality of supervision include focus (of project implementation) on development impact; supervision of fiduciary and safeguard or environmental and social aspects (when applicable); adequacy of supervision inputs and processes (including missions, as well as location and availability of key staff); candor and quality of performance reporting; and the World Bank's role in ensuring adequate transition arrangements (for regular operation of supported activities after loan/credit closing).

[15] Quality at entry ratings are assigned at project closing, not at project appraisal.

[16] A regression of average project outcome ratings on project development objective restructuring, country capacity measured by the Human Capital Index, lending volume, and days between project closing and ICRR completion showed that the coefficient of the dummy variable for projects that closed in FY20 remained large and significant. The *RAP* applied several different specifications, but—although the regressors cited above were all statistically significant in explaining average outcome ratings—the positive effect for FY20 remained largely unexplained and was not absorbed by a differential impact of the regressors for FY20.

[17] First-time projects and successor projects with limited novelty are more frequently used in fragile and conflict-affected situation countries.

[18] The hypothesis is that outcome ratings may be explained by the type of development outcomes that the project aims to achieve and, specifically, that some outcome types may be inherently more or less likely to be achieved. Outcome types offer a new way to classify projects and represent characteristics that are comparable across projects of different Practice Groups, GPs, and Regions.

[19] This may be due to the fact that these objectives also pursue another outcome type, which would be the one with one or more associated indicators. In any case, it is still hard to justify that objectives with clear institutional strengthening outcomes do not have any indicator to measure that outcome.

[20] This category of weak evidence included projects where the ICRR outcome sections collated narratives from various sections of the ICR or reported observations from interviews with project task team leaders.

[21] A factor to consider is that projects often include multiple indicators to measure individual objectives, and the ICRR evaluator (who assigns the objective-level efficacy rating) is left to

decide how to weigh the various pieces of evidence, based only on what is presented in the ICR, when assigning ratings.

[22] The objective-level efficacy rating is assigned by the ICRR, not the ICR. Therefore, if the indicator table in the ICR lists a target as exceeded, and the ICRR evaluator finds that evidence provided indicates the target was not in fact exceeded, the rating in the ICRR may be lower than the rating implied by the ICR narrative.

[23] Many indicators had a zero baseline. Out of 428 indicators in projects with negligible project efficacy ratings, 226 (53 percent) had a zero baseline; out of 217 indicators in projects with high or substantial project efficacy ratings, 87 (40 percent) had a zero baseline. For example, about 10 percent of project development objective indicators measured the number of direct project beneficiaries. All but 13 of the 60 indicators measuring beneficiaries had baselines of zero (5 of them also had no target, neither original nor revised). In these cases, the information on beneficiaries measures the size of the project's activities but does not measure how well the project met development needs. A similar observation holds regarding the proportion of female beneficiaries: This indicator was often set to zero at baseline, rather than being expressed as a gender gap.

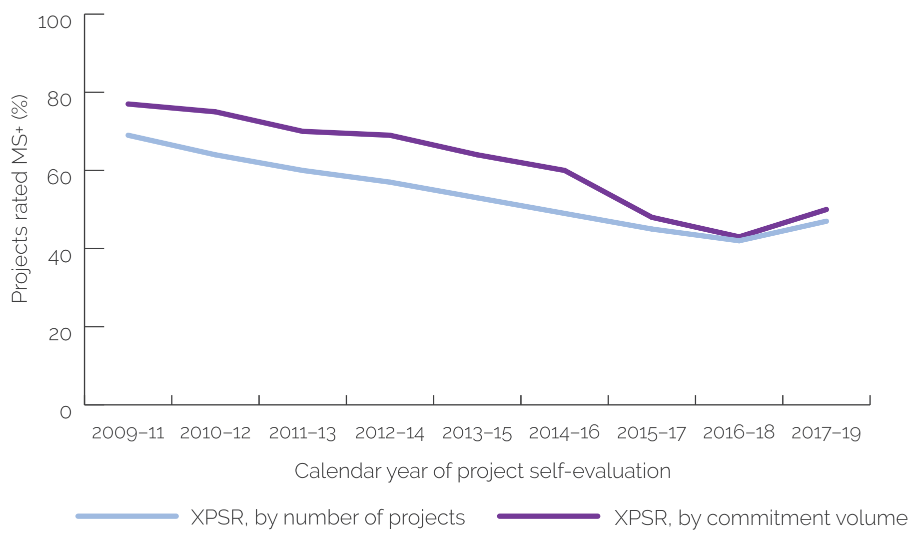
# 3. International Finance Corporation Results and Performance

IFC investments saw a recent improvement in ratings after years of decline. This uptick in ratings was possibly driven by changes in IFC's portfolio composition and its recent efforts to strengthen XPSR work. More generally, various factors, such as project outcome types, are likely to have influenced IFC's longer-term investment ratings trend.

## Recent Investment Ratings Improvements

IFC's overall development outcome ratings for investment projects improved recently after years of decline. IFC ratings are determined in XPSRs in a process that first involves IFC's self-evaluation of IFC investments. The self-evaluation is then validated by IEG using an EvNote.[1] For projects that close before sampling for evaluation, IEG prepares a Project Evaluation Summary (PES) in lieu of an XPSR.[2] The XPSR examines a project's business, economic, environmental, and social performance and its contribution to private sector development in the country. As shown in figure 3.1, IFC's development outcome ratings were declining in XPSRs until CY16–18, when 42 percent of projects were rated mostly successful or better (MS+). When measured annually, the overall development outcome success rate was its lowest in CY17, at 41 percent. This decline coincided with IFC's expansion to new clients, difficult countries, and more complex projects. In addition, these sets of projects evaluated during the period of decline included those projects approved during the global financial crisis of 2008–10.[3] However, IFC's investment performance has improved since 2018. IFC investment projects' development outcome ratings reversed their declining trend from the past 10 years, improving to 47 percent MS+ (by the number of projects) and 50 percent MS+ (by the net commitment volume) on a three-year rolling basis in CY17–19 (figure 3.1). IFC's annual overall development outcome success rate improved by 18 percentage points to 60 percent in CY19. In addition, as shown in figure 3.2, the share of successful projects increased, and the share of highly unsuccessful projects shrunk. It should be noted that the preliminary trends (from a still incomplete sample of validated projects) for CY20 indicate that the overall development outcome success rate might be lower than it was in CY19, so the recent rating uptick may not be sustained. Evaluations of projects affected by COVID-19 have not yet been included in the CY19 cohort.

**Figure 3.1.** IFC Investment Project Development Outcome Ratings

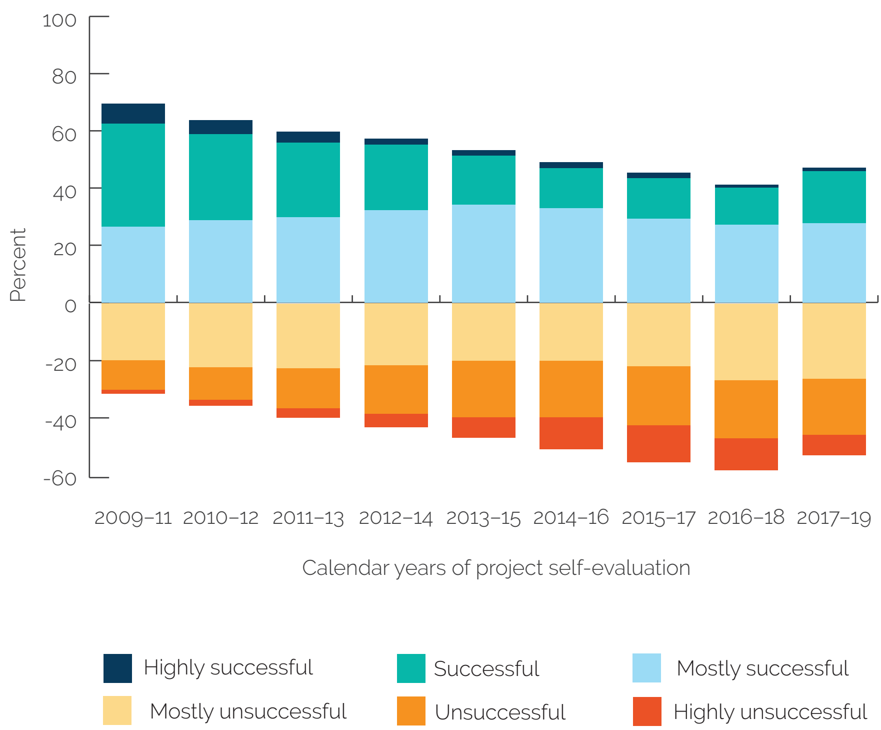



*Source:* Independent Evaluation Group.

*Note:* IFC = International Finance Corporation; MS+ = mostly successful or better; XPSR = Expanded Project Supervision Report.

Three out of IFC's four industry groups show a similar trend of improved ratings (figure 3.3). Projects in three of IFC's four industry groups—the Financial Institutions Group, Infrastructure, and Manufacturing, Agribusiness, and Services—had their lowest development outcome success rates of 45, 39, and 44 percent for CY16–18, before improving to 50, 48, and 47 percent in CY17–19. The fourth industry group—Disruptive Technologies and Funds—had the lowest development outcome success rate of 0 percent in CY15–17 (although this was a very small sample of projects) before substantially improving to 23 percent in CY17–19 (figure 3.3). On an annual basis, Financial Institutions Group, Infrastructure, and Manufacturing, Agribusiness, and Services showed increasing trends from CY18 to CY19, with success rates (MS+) of 57, 68, and 57 percent. A decomposition analysis for IFC projects confirms these findings. Most industry groups, except Disruptive Technologies and Funds, increased their shares between CY18 and CY19, but it was the increased success rate for all the industry groups that contributed most to the uptick (figure 3.4). For the whole decomposition analysis, please see appendix D.

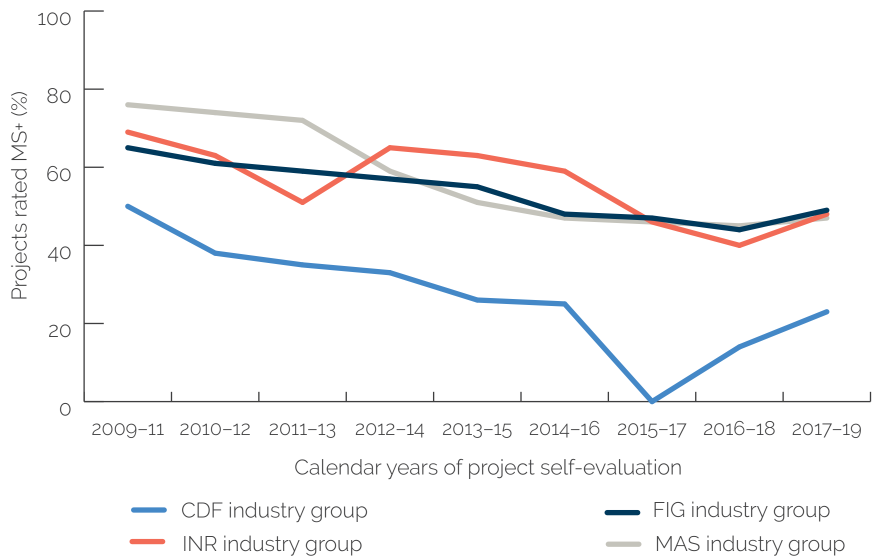
**Figure 3.2.** Distribution of IFC Development Outcome Ratings




*Source:* Independent Evaluation Group.

*Note:* IFC = International Finance Corporation.

**Figure 3.3.** IFC's Investment Project Development Outcome Ratings by Industry Group

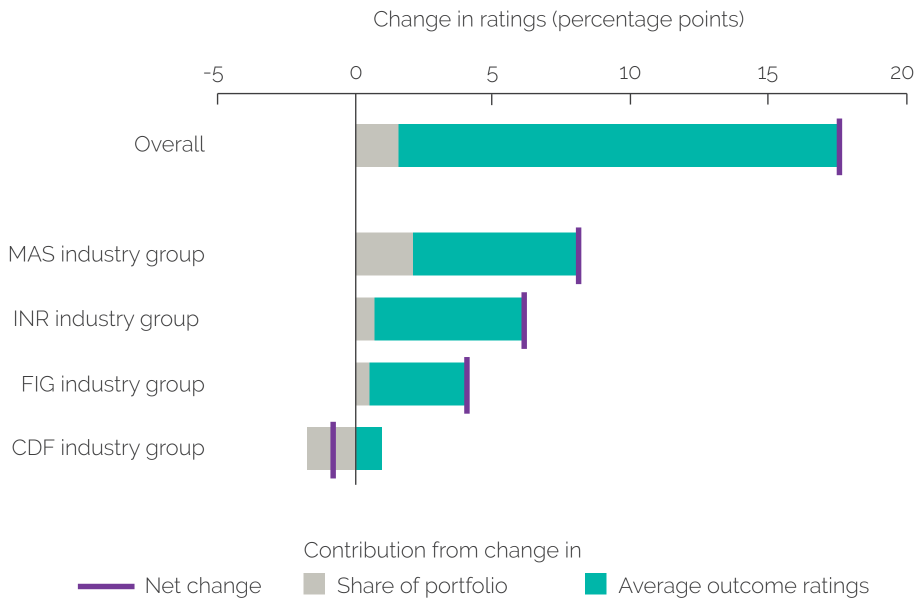



*Source:* Independent Evaluation Group.

*Note:* CDF = Disruptive Technologies and Funds; FIG = Financial Institutions Group; IFC = International Finance Corporation; INR = Infrastructure; MAS = Manufacturing, Agribusiness, and Services; MS+ = moderately satisfactory or above.

Well-known factors that influence investment performance likely continued to influence IFC's recent ratings trends. Previous *RAPs* identified factors that have contributed to declining trends over the years. The *RAP 2020*, in particular, identified internal work quality issues, external risks, and broader market trends as factors that drive IFC's investment project ratings. Meanwhile, a 2017 joint IFC-IEG study on IFC's work quality on investment projects identified staffing, incentives, diffused accountability, IFC's organizational culture, and IFC's focus on volume targets over development results as the main factors affecting IFC's investment work quality. IEG's review of project validations found that sponsor risks, market risks, country risks, and transaction structuring factors are clearly associated with IFC's investment project performance. IFC's efforts to partner with repeat clients are meant to address the sponsor risk (see box 3.1.) Broader market trends, including a weaker pool of viable investment projects, have exposed IFC to higher risks, which are hard to completely mitigate and may influence ratings. This *RAP* did not carry out an in-depth assessment on the impact of these well-known factors on the ratings increase from CY19, but it is likely they played a role.[4]

## Figure 3.4. Decomposition Analysis of the Increase in IFC Development Outcome Ratings by Industry Group (change between Calendar Years 2018 and 2019)




*Source:* Independent Evaluation Group.

*Note:* CDF = Disruptive Technologies and Funds; FIG = Financial Institutions Group; IFC = International Finance Corporation; INR = Infrastructure; MAS = Manufacturing, Agribusiness, and Services.

### Box 3.1. Working with Repeat Clients

Projects that worked with repeat clients showed higher ratings. Previous *Results and Performance of the World Bank Group* reports, Independent Evaluation Group (IEG) evaluations, and synthesis analyses showed that repeat clients reduced sponsor selection risks and had a positive impact on project performance and development outcome ratings. According to an IEG evaluation on the International Finance Corporation's (IFC) client engagement, 70 percent of projects with repeat clients were rated mostly successful or better, compared with 49 percent for IFC's one-off clients (World Bank 2018a). Higher performance mainly reflects a selection effect, since IFC is more likely to continue to engage with relatively stronger clients with a track record of implementing projects successfully and better fit with IFC's strategic focuses.

*(continued)*

### Box 3.1. Working with Repeat Clients (cont.)

Working with repeat clients would provide opportunities for IFC to leverage these relationships to enhance project ambition and maintain good performance. IEG analyses show that IFC's support and contribution to repeat clients could evolve during the engagement and exhibit a life cycle in which clients at some point outgrow the need for IFC's services. The analyses show that there were cases where IFC's first intervention with a client introduced new or better standards in corporate governance, environmental and social performance, or other areas, whereas the second intervention with the same client developed knowledge and innovation or expanded investments to different countries, unknown territories, higher-risk areas, or higher-risk markets, such as those in International Development Association countries and countries with fragile and conflict-affected situations. The need for IFC to maintain additionality throughout this life cycle in client relationships has implications for IFC's future selection and segmentation of clients. However, according to the IEG reports, there was no established pattern of these incremental additionalities, and ensuring additionality in follow-on projects remains a challenge. The IEG evaluation on IFC's client engagement called for strengthening criteria for additionality for strategic clients, including for the justification of incremental additionality (World Bank 2018a). IEG evaluations also emphasize the need for IFC to constantly renew and expand its client pool. For example, given the limited number of clients in fragile and conflict-affected situations that meet IFC standards, a recent IEG evaluation on IFC's and the Multilateral Investment Guarantee Agency's support for private investment in fragile and conflict-affected situations indicates that to increase business in those countries, IFC needs to broaden its client base to reach and build the capacity of local and regional private investments (World Bank, forthcoming).

*Source:* Independent Evaluation Group; World Bank, forthcoming.

IFC's efforts to address these factors may have paid off in the recent ratings improvements. In the past few years, IFC created the Economics and Private Sector Development Vice Presidential Unit to strengthen its project and macroeconomic analyses, launched the AIMM framework, and strengthened the Accountability and Decision-Making Framework. IFC's management has improved the quality of self-evaluations, including investment projects'

XPSRs and advisory services projects' Project Completion Reports (PCRs), and engaged more proactively in reviewing IEG's EvNotes and PESs. These combined efforts led to IFC management providing targeted, expert advice to improve project analyses and help articulate project outcomes, including a project's development impact, in the XPSRs and PCRs. These efforts also facilitate the processing and effective management of XPSRs, PESs, and PCRs. It is possible, but not confirmed, that some of these efforts may be reflected in the recent ratings uptick. For example, after management's push to improve XPSRs, the share of XPSRs nominated as best practices increased from 12 percent in 2016, 11 percent in 2017, and 10 percent in 2018 to 20 percent in 2019.

These efforts increased the dialogue between IFC and IEG on project self-evaluations and reduced IEG-IFC ratings variance. As a result, the share of EvNote and PESs to which IFC provided comments increased from 74 to 84 percent from CY16 to CY19. More important, the quality of the responses and information provided by clients increased substantially, in IEG's view. These improvements prompted additional dialogue between IFC and IEG, which resulted in enhanced learning on both sides and reduced ratings variance between the two groups. During the XPSR evaluation process, IFC's self-rating and IEG's validation rating often differ; this is referred to as ratings variance. The increased dialogue between IFC and IEG, enhanced by joint training on self-evaluation and support from experts, built a common understanding by both groups of self-evaluation guidelines, facilitated the collection of quality evidence, and improved the objective assessments of project performance. All of these efforts helped reduce the ratings variance from 31 percent in CY17 to 8 percent in CY19.

Recent IEG Sector Highlights and validation exercises have shed light on what factors contributed to the recent ratings improvement. Financial Institutions Group's Sector Highlights, carried out in FY21, identified several high-probability (frequency) and high-impact (severity) risks that were common in low development outcome ratings. These risks include business, economic, foreign exchange, management, and corporate governance risks. Manufacturing, Agribusiness, and Services's Sector Highlights, carried out in FY20 and focused on the agribusiness portfolio, took a similar approach and discovered that (i) project, country, or macro factors influenced projects from previous years, but diminished as factors for more recent projects as the impact of the 2008–10 financial crisis subsided; (ii) market risks related to prices and competition were associated with negative outcomes in agribusiness projects, more than in other Manufacturing, Agribusiness, and Services projects; and (iii) sponsor and management quality were important factors for both successful and failed projects.[5] IEG's validations of Infrastructure's CY19 projects found that many factors positively influenced ratings, including transaction structuring aspects, a sponsor or company's management capacity, a company's stronger competitiveness in the market, and a favorable political or regulatory environment. Meanwhile, changing dynamics in the market and macroeconomic situations, such as currency depreciation, negatively influenced ratings.

In CY19, IFC's subsector composition had fewer poorly performing clients, which contributed to the improvement of overall development outcome ratings. In Infrastructure, there were fewer platform companies in the power sector, junior miners in the mining sector, and nonmobile telecom clients in the telecom, media, and technology sectors. These types of clients tend to have lower ratings than other clients. In Manufacturing, Agribusiness, and Services, fewer retail, tourism, construction, and real estate projects (whose performance declined in CY19), combined with more agribusiness and manufacturing projects (whose performance improved in CY19), contributed to IFC's aggregated improved ratings.

For the Financial Institutions Group, the lower share of greenfield projects and improved ratings of projects in Europe and Central Asia contributed to positive results in CY19 compared with previous years. Greenfield projects are projects that finance new ventures and activities.[6] Historically, greenfield projects have lower development outcome ratings than mature projects because they take more time to deliver results and many things can go wrong, although careful structuring and close supervision can help reduce or mitigate the risks. Another factor behind Financial Institutions Group's improved performance was a slight improvement in recent years of development outcome ratings in Europe and Central Asia. However, the increasingly unstable economic environments since 2019, which include political instability and geopolitical conflicts, could affect investor sentiments and market performance.

## Longer-Term Investment Rating Trends

IFC's ratings have generally declined over the past 10 years, a trend that differs from the ratings uptick in CY19. This section explores some potential influences on these longer-term ratings trends. This analysis shows that characteristics such as market-level outcome types and a project's development potential may contribute to changes in ratings. This chapter's analysis relies on a combination of AIMM, XPSR, and EvNote data, which box 3.2 describes.

### Box 3.2. Analysis of IFC's Outcome Types and Outcome Potential

This chapter uses the Anticipated Impact Measurement and Monitoring (AIMM), Expanded Project Supervision Report (XPSR), and Evaluation Note (EvNote) frameworks for project performance data. The *Results and Performance of the World Bank Group* (*RAP*) report's objective was to use these data to analyze the impact on ratings of International Finance Corporation (IFC) outcome types and outcome potential. As in the chapter 2 analysis, outcome types refer to the outcomes that project activities pursue. The *RAP* defined 13 outcome types for IFC projects. Outcome potential, which is unique to IFC, refers to the magnitude of the development challenge that a project is going to address in a given country or sector and the intensity or extent to which these efforts contribute to those development challenges. XPSRs and EvNotes provide evaluation ratings data for projects, whereas AIMM provides additional data on outcome types and outcome potential, which are not measured by XPSRs and EvNotes. In other words, we use AIMM data to supplement XPSR and EvNote data. However, IFC only introduced the AIMM framework in 2017. Therefore, this *RAP* had to review the projects to which IFC retroactively applied AIMM data ("backfilled projects"), whose approval predates the introduction of the framework, and fill in missing XPSR and EvNote ratings. The Independent Evaluation Group has not yet evaluated or validated the projects that applied AIMM at approval; therefore, these projects were excluded from this *RAP*'s sample of IFC projects. As a result of these various frameworks and the need to rely on projects with backfilled AIMM data, IFC's project results and the Independent Evaluation Group's findings from the outcome type and outcome potential analyses should be interpreted cautiously.

*(continued)*

## Box 3.2. Analysis of IFC's Outcome Types and Outcome Potential (cont.)

**Methodology for outcome type analysis.** The outcome types defined for this exercise capture the type of change envisioned by project objectives. For IFC, we defined an outcome typology that includes 13 outcome types based on the AIMM sector framework. We analyzed outcome types for all projects that had AIMM monitoring results based on backfilled AIMM and in XPSRs validated between fiscal year (FY)12 and the first half of FY20. We carried out our own assessment on outcome claim achievement by reviewing XPSRs and EvNotes, complemented with AIMM intensity and movement ratings for outcome claims at the approval and monitoring stages. Appendixes A and B provide further details.

**Methodology for outcome potential analysis.** To assess the relationship between projects' outcome potential and their XPSR rating, we used AIMM potential ratings (both at the approval and monitoring stages) for AIMM's two dimensions—project outcome and market outcome, with XPSR's ratings for economic sustainability and private sector development, two of XPSR's four dimensions. Although AIMM and XPSRs assess similar elements for their corresponding dimensions (such as project outcome and economic sustainability or market outcome and private sector development), the ways these outcomes are assessed differ between AIMM and XPSR because they serve different purposes: AIMM is an ex ante and monitoring framework and XPSR is an evaluation framework. To build the cohort, we used the same projects used for the outcome type analysis. Appendix A provides additional details.

The outcome type analysis and the outcome potential analysis are based on the early-stage implementation of AIMM framework and projects with backfilled AIMM data. Therefore, as mentioned above, their findings and implications should be interpreted cautiously.

*Source:* Independent Evaluation Group.

*Note:* IFC = International Finance Corporation.

According to AIMM's analytical framework, IFC's project outcomes fall into two broad categories—project-level claims and market-level claims. Project-level claims, or outcomes, are defined as a project's direct and indirect effect on stakeholders, the economy, and the environment. Market-level claims are derived effects, defined as a project's ability to catalyze systemic changes beyond those brought about by the project itself. Overall, all of the reviewed IFC projects had project-level claims and 86 percent had market-level claims. Furthermore, IFC projects focused on two outcome types the most, with 68 percent pursuing the project-level claim "improved access to goods and services," and 53 percent pursuing the market-level claim "increased market competitiveness."

Projects are less likely to achieve market-level claims, particularly competitiveness in the market, than project-level claims.[7] The "competitiveness," "integration," and "sustainability" market-level outcome types have the lowest achievement rates of all outcome types, at 36 percent, 43 percent, and 38 percent (table 3.1). Among the project-level outcome claims, access to goods and services shows a relatively low success rate. Market-level outcome types also have a larger share of downgraded AIMM claim ratings than project-level outcome types.

These results show that it is more difficult for IFC to achieve and measure market-level outcomes than project-level outcomes. This is because the success of market-level outcomes depends on the broader market environment and external factors such as market changes and actions by external actors, including government officials or private companies. Also, measuring market-level outcomes is challenging because of the long-term time horizons for outcomes to materialize, the challenge of attributing market-level results to IFC-supported projects, fewer good indicators to measure projects' contribution with certainty, and the minimal impact that an individual IFC project can have on the broader market.[8, 9] By contrast, project-level outcomes have shorter time horizons and often provide goods, services, financing, or infrastructure, all of which IFC and its counterparts have more control over achieving. Also, these outcome types tend to have more accurate and attributable evidence to use for monitoring performance. That is not to say that project-level outcomes are not challenging in their own way. For example, "access to goods and services for MSMEs [micro, small, and medium enterprises]" has a relatively lower achievement rate and larger variance in change ratings. This outcome type requires expanded lending to micro, small, and medium enterprises, enabling them to borrow from financial institutions, which is not entirely within the control of the project, either. Also, project-level outcomes may be just as risky as market-level outcomes. Our methodologies used variance, or the change in outcome-claim level AIMM ratings between the approval and monitoring stages, as a measure of risk and uncertainty. "Access to goods and services," for example, is a project-level outcome that shows a relatively large variance in claim rating changes.

## Table 3.1. Outcome Type Performance and Performance Changes from the Approval Stage to the Monitoring Stage, 2012–20

| Claim Type | Outcome Claims (% achieved) | Outcome Claims (no.) | Change in Claim Rating (%) −2 | −1 | 0 | +1 | +2 |
|---|---|---|---|---|---|---|---|
| Project-level claim | | | | | | | |
| 1.1 Access to goods and services (1.1.1–1.1.3) | 51 | 193 | 4 | 22 | 63 | 10 | 1 |
| 1.1.1 Access to goods and services (MSME) | 51 | 97 | 5 | 27 | 57 | 11 | 0 |
| 1.1.2 Access to goods and services (female) | 71 | 14 | 8 | 8 | 62 | 15 | 8 |
| 1.1.3 Access to goods and services (customers) | 53 | 97 | 2 | 15 | 73 | 10 | 0 |
| 1.2 Quality/affordability of goods and services | 63 | 16 | 13 | 13 | 67 | 7 | 0 |
| 2.1 Suppliers/distributors reached | 68 | 25 | 0 | 18 | 73 | 9 | 0 |
| 2.3 Improved sales/profitability of suppliers/distributors | 66 | 29 | 0 | 16 | 76 | 8 | 0 |
| 3.1 Increased employment | 57 | 56 | 2 | 16 | 73 | 9 | 0 |
| 6.2 GHG reduction | 70 | 30 | 0 | 23 | 77 | 0 | 0 |
| Market-level claim | | | | | | | |
| 9. Competitiveness in the market | 36 | 126 | 10 | 19 | 67 | 5 | 0 |
| 10. Resilience in the market | 63 | 24 | 4 | 26 | 70 | 0 | 0 |
| 11. Integration in the market | 43 | 28 | 11 | 11 | 75 | 4 | 0 |
| 12. Inclusiveness in the market | 69 | 16 | 6 | 38 | 44 | 13 | 0 |
| 13. Sustainability in the market | 38 | 13 | 0 | 46 | 54 | 0 | 0 |

*Sources:* Independent Evaluation Group and International Finance Corporation.

*Note:* (i) Of project-level outcome claims considered not achieved, 25 percent were considered not achieved because their results could not be verified. Of market-level outcome claims considered not achieved, 57 percent were considered not achieved because their results could not be verified. The relatively high percentage of nonverified market-level outcomes was partly due to the fact that many of those market outcomes were identified retroactively with backfilled Anticipated Impact Measurement and Monitoring frameworks. Monitoring of outcome results, particularly of market-level outcome, may improve under the Anticipated Impact Measurement and Monitoring framework. (ii) "Change in Claim Rating" means step changes of intensity (project claims) or movement (market claims) ratings for claims from ex ante to monitoring. GHG = greenhouse gas; MSME = micro, small, and medium enterprise.

Projects that had market-level claims at approval did not necessarily end up with lower project-level ratings, partly because of different evaluative approaches. We assessed how project development outcomes that intended to achieve a market impact were rated by XPSR.[10] The results indicate that projects with market-level intentions had XPSR ratings similar to those of projects without a market-level intention (table 3.2). Moreover, the variability of XPSR ratings for projects with market intentions does not seem to be particularly high compared with XPSR ratings for projects without market intentions (figure 3.5). One reason for this is the XPSR evaluation approach. For example, XPSRs account for the achievement of not only expected market outcome claims but also various other elements such as (i) the performance of relevant competitors and industry benchmarks in the prevailing macroeconomic environment; (ii) other unintended outcomes in private sector development; and (iii) business performance, economic sustainability, and environmental and social effects. Therefore, if a project does not achieve a market-level claim, it does not automatically mean that the project's development outcome will be low in XPSR. At the same time, if a project achieves a transformational market impact, it may be properly reflected by a higher private sector development rating. Refer to box 3.3 for some examples of projects with substantial market impact.

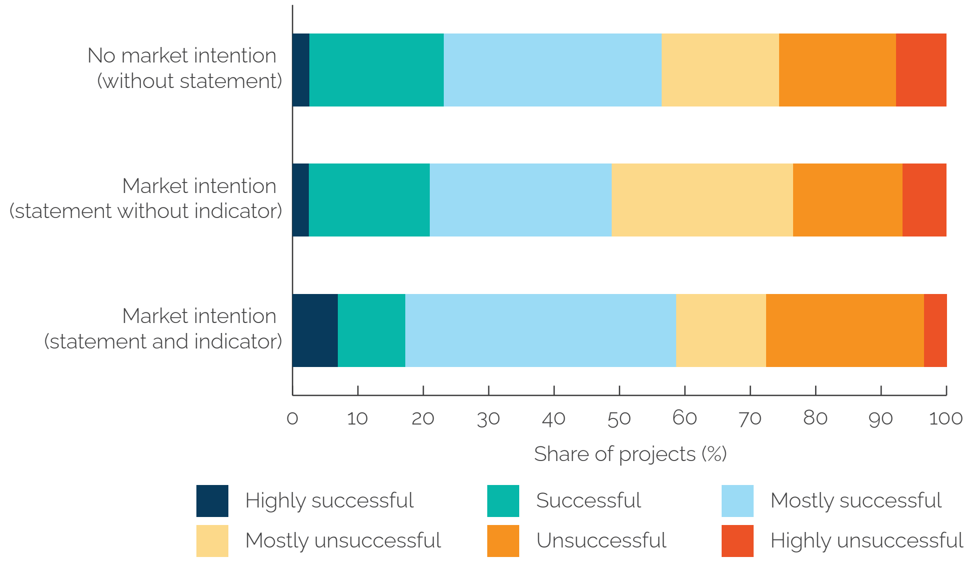
## Table 3.2. AIMM and XPSR Ratings of Projects with Market-Level Outcome Intentions

| Market Intention | AIMM Ex Ante Score (avg.) | XPSR DO (avg.) | HS | S | MS | MU | U | HU |
|---|---|---|---|---|---|---|---|---|
| 1. No market intention (without statement) | 26.0 | 3.5 | 1 | 8 | 13 | 7 | 7 | 3 |
| 2. Market intention (statement without indicator) | 40.6 | 3.4 | 4 | 30 | 45 | 45 | 27 | 11 |
| 3. Market intention (statement and indicator) | 45.2 | 3.5 | 2 | 3 | 12 | 4 | 7 | 1 |

*Source:* Independent Evaluation Group.

*Note:* avg. = average; AIMM = Anticipated Impact Measurement and Monitoring; DO = development outcome; HS = highly successful; HU = highly unsuccessful; MS = mostly successful; MU = mostly unsuccessful; S = successful; U = unsuccessful; XPSR = Expanded Project Supervision Report..

## Figure 3.5. Distribution of XPSR Ratings by Project Market Intention




*Source:* Independent Evaluation Group.

*Note:* XPSR = Expanded Project Supervision Report.

## Box 3.3. Examples of IFC Investment Projects with Market Impacts

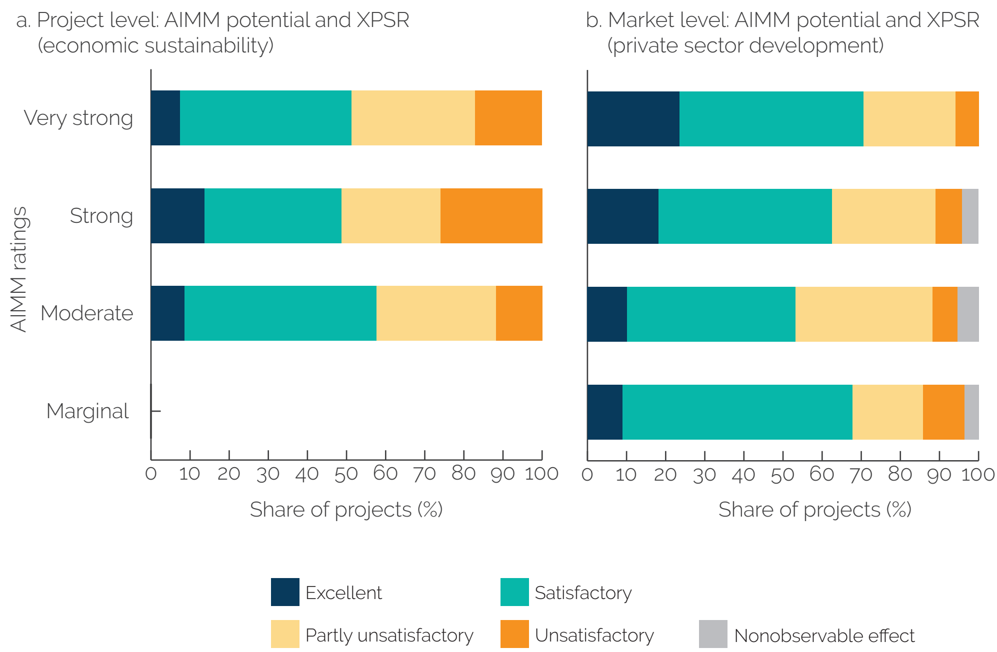
**Providing sustainable energy financing in Lebanon.** The International Finance Corporation provided $10 million to expand the sustainable energy finance (SEF) portfolio of a commercial bank in Lebanon. It also provided advisory services to the bank to assess risk and develop and market SEF products in Lebanon. As a result, the bank successfully expanded its SEF portfolio, which led to multiple repeat transactions with this bank and other financial institutions in the country and created a new market for climate change financing and SEF lending. All of this eventually led to the issuance of the first green bond in Lebanon and the Levant region.

**Developing the first mobile virtual network operator in Chile.** The International Finance Corporation provided a $11 million C loan in Chile to launch and operate a mobile virtual network operator (MVNO)—a wireless communications provider that does not own the wireless network infrastructure. The MVNO primarily targeted Chilean youth. The project was the country's first MVNO, thereby having a strong demonstration effect, leading to seven more MVNOs launching in Chile. The multiple MVNOs in the country increased competition among mobile services and helped reduce prices for consumers.

*Source:* Independent Evaluation Group.

Projects with high development potential did not have lower XPSR ratings. A high development potential means a higher magnitude of development challenges in a given country and a more intense IFC contribution toward these challenges. As figures 3.6 and 3.7 show, projects with high development potential are not accompanied by lower XPSR ratings or more variance, which undermines a common assumption that the more sophisticated or challenging outcomes associated with higher development potential would contribute to lower ratings. However, this finding should be interpreted cautiously, since it is based on an analysis of projects with backfilled AIMM data. The positive correlation between development potential and ratings is an interesting finding, but further analysis is required to determine what drives it. (For a comparison between the AIMM and the XPSR frameworks, see appendix A.)

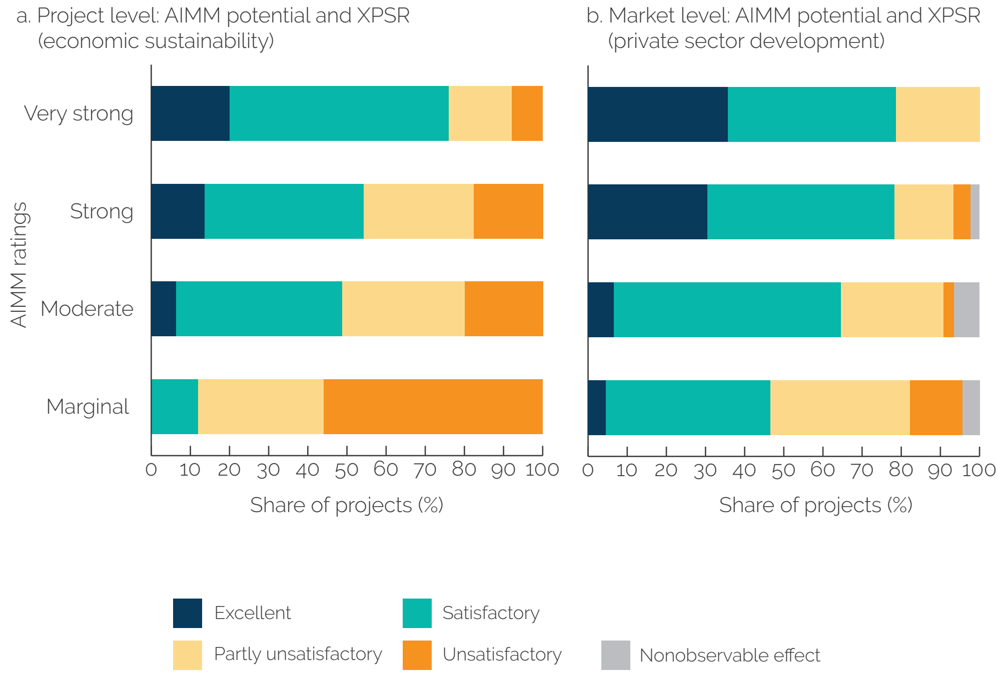
### Figure 3.6. XPSR Ratings by AIMM Development Potential Ratings at Approval




*Source:* Independent Evaluation Group.

*Note:* Panels show the percentage of projects with AIMM development potential ratings at approval, from marginal to very strong, with XPSR ratings in economic sustainability (panel a) and private sector development (panel b) from unsatisfactory to excellent. AIMM = Anticipated Impact Measurement and Monitoring; XPSR = Expanded Project Supervision Report.

## Figure 3.7. XPSR Ratings by AIMM Development Potential, Ratings at Monitoring



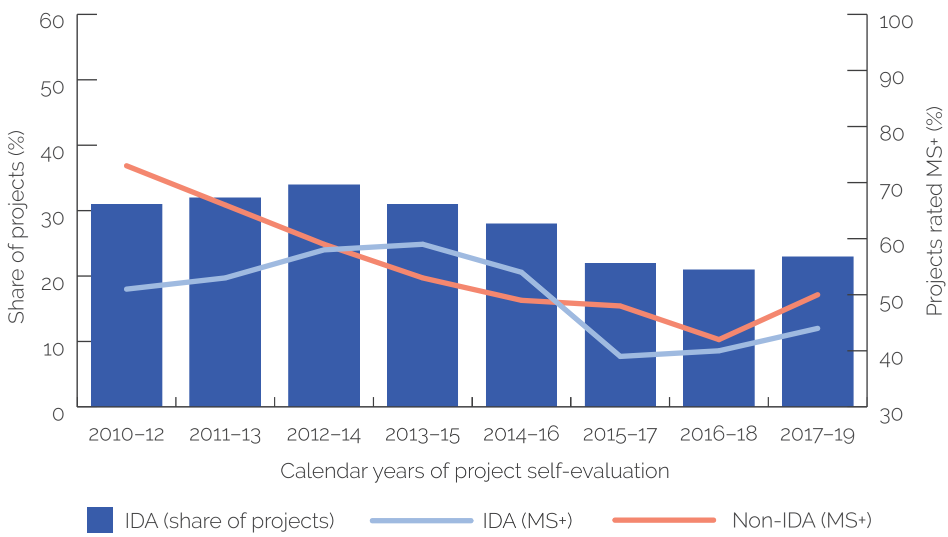

*Source:* Independent Evaluation Group.

*Note:* Figures show the percentage of projects with AIMM development potential ratings at monitoring, from marginal to very strong, with XPSR ratings in economic sustainability (panel a) and private sector development (panel b) from unsatisfactory to excellent. AIMM = Anticipated Impact Measurement and Monitoring; XPSR = Expanded Project Supervision Report.

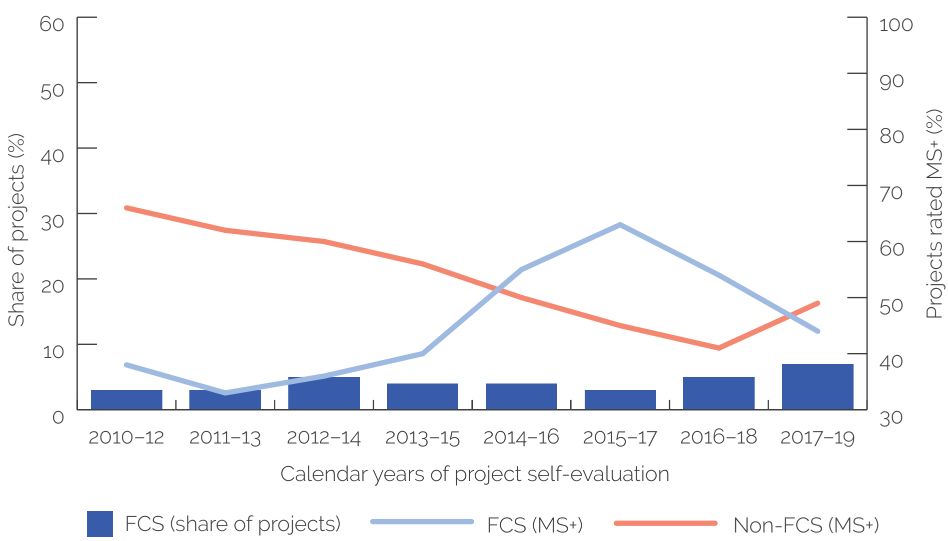
IFC projects that were aligned with prominent corporate priorities did not have consistently lower ratings than projects that did not. We looked at projects that addressed IFC's corporate priorities of climate change, IDA countries, FCS, and inclusive business, which includes gender priorities. Box 3.4 describes each of these priorities in more detail. Figure 3.8 shows the success rate (defined as having an overall XPSR development outcome rating of MS+) of projects with or without different corporate priorities. Since 2010–12, the performance of the two categories of projects has varied, with no obviously consistent pattern. In the most recent years, inclusive business projects had better results than projects that did not integrate inclusive business, whereas IDA projects, FCS projects, and climate change projects were less successful over a longer period, possibly contributing to IFC's longer-term ratings declines. That said, the performance of IDA projects and climate change projects has improved recently. The latter is noteworthy because of the various legal, technological, and market-related challenges that climate change projects face. As IFC projects increasingly embrace corporate priorities for the public good, it is critical to monitor their impact on project outcomes to ensure that proper mitigation measures are in place and that there are proper incentives for private enterprises to pursue these higher-level goals.

**Figure 3.8.** IFC Projects with Development Outcome Rated MS+, with and without Corporate Priority Objectives

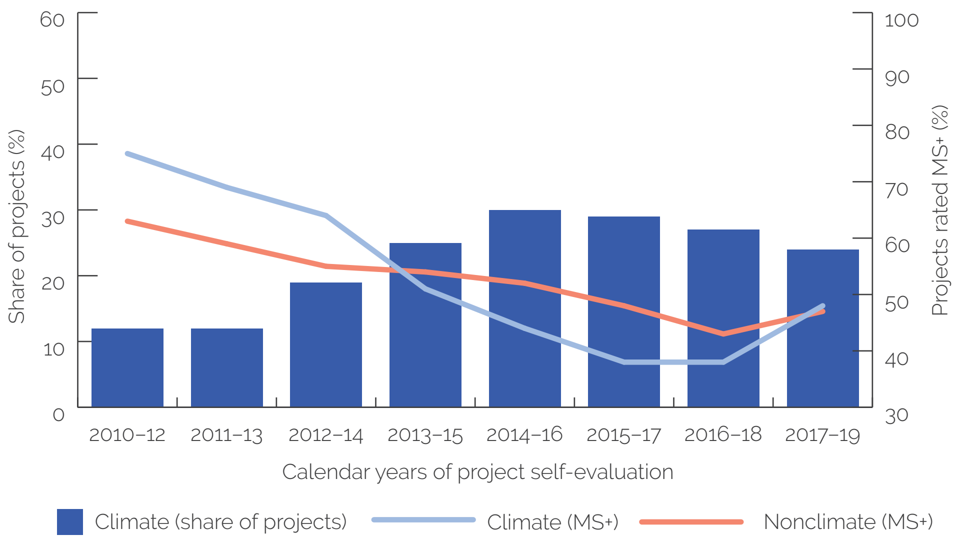
a. Performance of IDA projects



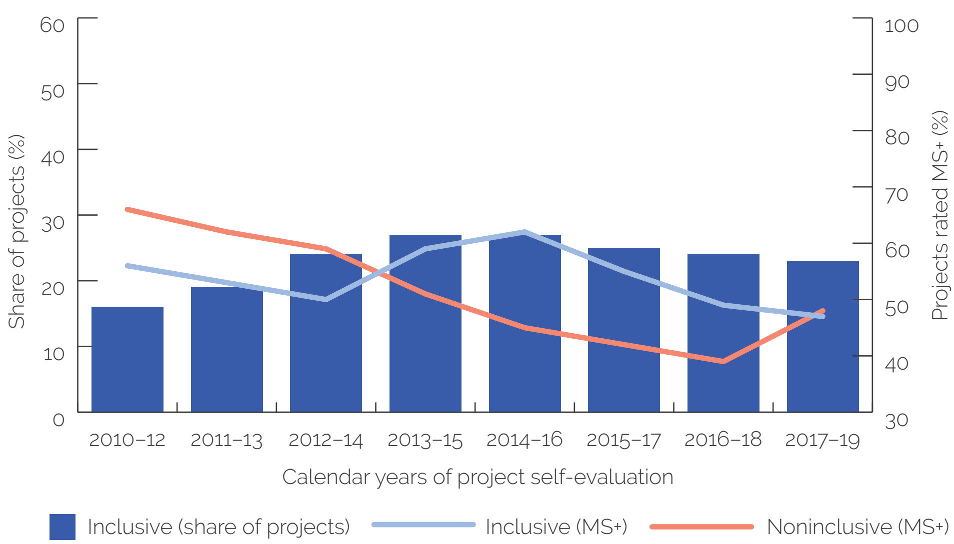

b. Performance of FCS projects



c. Performance of climate change projects




d. Performance of inclusive business projects




*Source:* Independent Evaluation Group.

*Note:* FCS = fragile and conflict-affected situation; IDA = International Development Association; IFC = International Finance Corporation; MS+ = mostly successful or above.

## Box 3.4. IFC's Corporate Priorities

**International Development Association (IDA) countries.** IDA countries have been a priority area for the International Finance Corporate (IFC). Most recently, as part of the 2018 capital increase package, IFC committed to delivering 40 percent of its overall business program to IDA and fragile and conflict-affected situation (FCS) countries, including 15–20 percent to low-income IDA and IDA FCS countries.

**Fragile and conflict-affected situations.** IFC has supported investments in FCS since 2009 and adopted an FCS strategy in 2012.

**Climate change.** The World Bank Group launched its first Climate Change Action Plan in 2016 and has released a second plan for 2021–25. In the second plan, IFC committed to aligning 100 percent of new operations to the Paris Agreement goals by 2025.

**Inclusive business, including gender.** In IFC, the inclusive business concept was initiated in 2010, together with the establishment of an inclusive business unit. Since then, IFC has supported inclusive business in a wide range of industries.[a]

**Methodologies.** To assess the relationship between projects that intend to address corporate priorities and their Expanded Project Supervision Report ratings, we classified all projects evaluated since the calendar year 2012 cycle based on the flag provided by the IFC team for climate change and inclusive business and IDA and FCS classification at the time of evaluation.[b]

*Source:* Independent Evaluation Group.

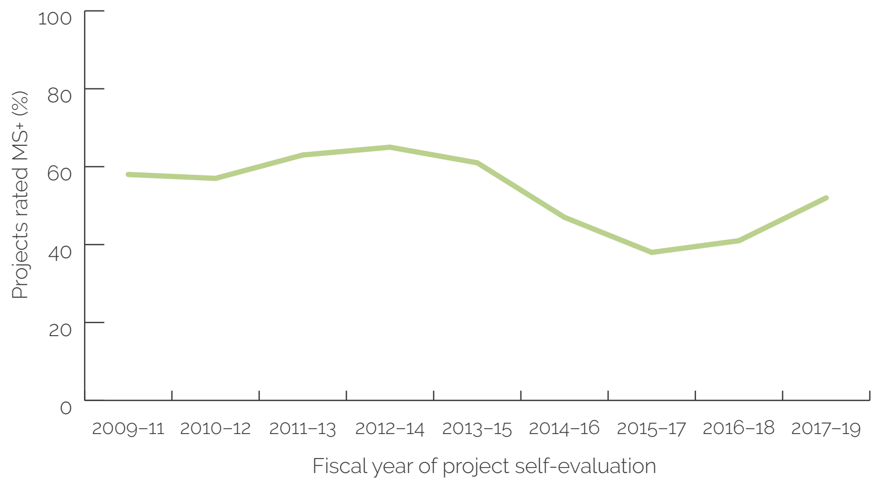
*Note:* a. Gender is another of IFC's corporate priorities. However, the *Results and Performance of the World Bank Group* report team decided to focus on inclusive business, which also includes gender aspects, rather than gender itself, because the Bank Group's gender strategy was only launched in 2016, so projects were not flagged for gender elements before that year.
b. Subject to the availability of data at the time of assessment. For country classification for IDA and FCS, historical country classification based on the Bank Group's classification was applied for evaluation year of projects. For example, for a project in the 2016 evaluation year, IDA and FCS country classifications of the Bank Group in 2016 were applied.

## Advisory Services

Development effectiveness ratings for IFC's advisory services projects continue to improve (figure 3.9). Development effectiveness ratings of MS+ fell to their lowest levels in FY15–17 at 38 percent, but they have been improving ever since, reaching 41 percent in FY16–18 and 52 percent in FY17–19.

When calculated by advisory projects' total funding amount rather than the number of projects, development effectiveness ratings declined from 70 percent MS+ in FY12–14 to 33 percent in FY15–17, before increasing to 51 percent in FY17–19.

Figure 3.9. IFC Advisory Projects' Development Effectiveness Ratings




*Source:* Independent Evaluation Group.

*Note:* IFC = International Finance Corporation; MS+ = mostly successful or better.

Several factors influence the development effectiveness ratings for IFC's advisory projects. As reported in previous *RAPs*, large project sizes, longer project durations, and team leader changes have a statistically significant negative association with project success. Other factors that influence IFC's advisory services ratings include the client's commitment, robust project M&E, and IFC's flexible and proactive supervision. Moreover, this *RAP*'s analysis of recent evaluations revealed two additional aspects that have improved ratings, including improvements to IFC's work quality and the use of improved evidence to monitor development results.

IFC has taken actions to improve advisory services projects, possibly leading to better ratings. These actions were in response to findings from a joint IFC-IEG work quality study for advisory projects from 2017. These actions have improved IFC's annual work quality ratings since FY18 in both project preparation and design and project implementation and supervision, with the latter improving from 47 percent in FY17 to 74 percent in FY19. The effect of IFC's improved work quality at project implementation and supervision, particularly management's enhanced review of project status and the introduction of internal mechanisms to increase management's accountability at all project stages, has likely contributed to the overall development effectiveness ratings increase.

IFC's PCRs have shown improved M&E and use of evidence, which likely contributed to improved development effectiveness ratings. IFC's PCRs are self-evaluation documents for closed IFC projects. The improved ratings were particularly notable for development effectiveness and, most significantly, outcome ratings. For overall development effectiveness and outcomes, the share of PCRs that used quality evidence to a "sufficient extent" and a "great extent" increased from 62 and 46 percent in 2016 to 70 percent for both categories in 2019. The improved evidence base may have also helped reduce the rating variance in PCRs, where the difference between IFC and IEG ratings decreased from 41 percent in 2016 to 13 percent in 2019.

[1] An Expanded Project Supervision Report is a self-evaluation report prepared by the International Finance Corporation (IFC) project team and is typically prepared five to seven years after the project was approved by the Board. The EvNote (Evaluation Note) is a validation report of the Expanded Project Supervision Report prepared by the Independent Evaluation Group (IEG).

[2] In recent years, Project Evaluation Summaries account for about 25 percent of the projects selected for evaluation.

[3] Among the projects approved during the global financial crisis in 2008–10, 83 percent were evaluated in the 2013–15 cohort.

[4] IEG's review of the calendar year 2019 project cohort by a machine learning framework shows that sponsor risk, country risk, and market or industry risk are factors associated with projects' success or underperformance.

[5] The Manufacturing, Agribusiness, and Services Sector Highlights were originally prepared for 2014–18; additional analysis was carried out for the calendar year 2019 projects for the purpose of this *Results and Performance of the World Bank Group* report.

[6] The IFC Business Glossary defines *greenfield* as a "project where no institution, asset or operation currently exists."

[7] This assessment uses two metrics to assess the relationship between IFC's outcome types and project performance, including (i) the achievement of outcome claims, referred to as outcome claim achieved, at the time of evaluation, and (ii) the change of Anticipated Impact Measurement and Monitoring ratings for outcome claims between the project's approval and monitoring stages.

[8] In Anticipated Impact Measurement and Monitoring framework, IFC's assertions explicitly focus on IFC contribution to market changes, rather than attribution.

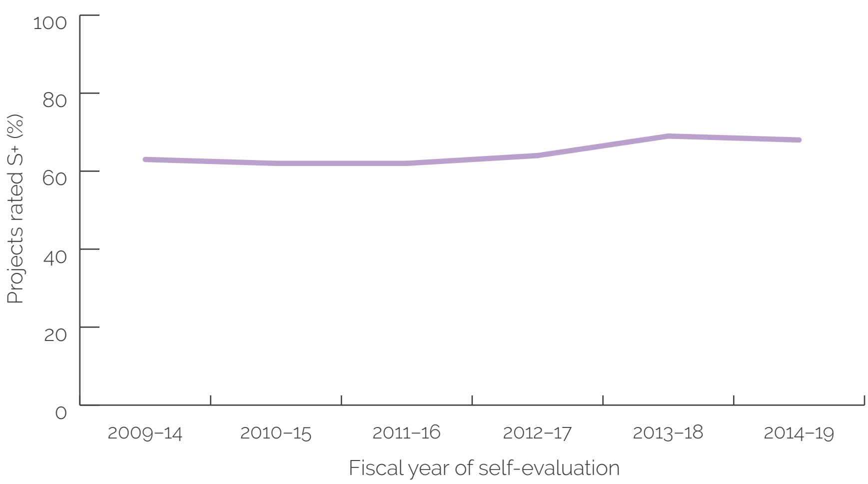
[9] IEG's thematic evaluation *Creating Markets to Leverage the Private Sector for Sustainable Development and Growth* (World Bank 2019a) reinforces this view by emphasizing the critical role an enabling environment plays in creating markets and calling for strengthened monitoring and evaluation systems for market creation projects.

[10] To classify projects with market outcome intentions, the *Results and Performance of the World Bank Group* report team divided projects into three groups: (i) projects without market outcome claims, (ii) projects with market outcome claims but without indicators, and (iii) projects with market outcome claims with indicators, the last group having greatest intent to achieve market impact.

# 4. Multilateral Investment Guarantee Agency Results and Performance

MIGA's project development outcome ratings have been increasing over the past 10 years. More specifically, MIGA's development outcome ratings increased from 62 percent S+ in FY11–16 to 68 percent S+ in FY14–19 (figure 4.1). When calculating ratings by gross issuance amounts, MIGA's development outcome ratings increased from 59 to 74 percent S+ over the same time frame.

Figure 4.1. MIGA Project Development Outcome Ratings

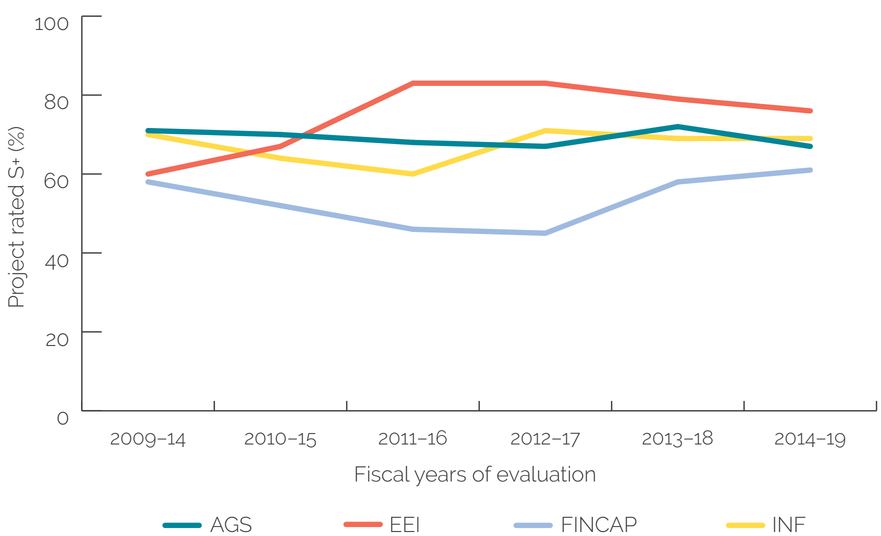



*Source:* Independent Evaluation Group.

*Note:* MIGA = Multilateral Investment Guarantee Agency; S+ = satisfactory or above.

MIGA's financial sector had the lowest performance, although this has been converging with the performance of other sectors. Among MIGA's four sectors, the Energy and Extractive Industries sector had the highest success rate, with 75 percent of development outcome ratings being S+, although this rate has declined recently. Part of the reason for this decline is that several power projects were unable to produce as much electricity as expected because of technical issues and increased competition with other energy sources. Meanwhile, MIGA supported several first-of-a-kind power projects in countries with the potential to achieve a demonstration effect. In MIGA's Infrastructure sector, there were several successful water and sanitation projects and urban transport projects, but telecom projects were less successful. MIGA's Agribusiness and General Services sector's performance was stable, although agribusiness projects, which are a majority of the sector's projects, had lower performance than other subsectors. MIGA's Finance and Capital Markets sector's performance improved in 2018 and 2019 after its portfolio shifted from shareholder loans from parent companies supporting their subsidiaries to capital optimization projects and state-owned enterprise projects. See figure 4.2 for sector ratings trends.

Figure 4.2. MIGA Project Development Outcome Ratings by Sector




*Source:* Independent Evaluation Group.

*Note:* AGS = Agribusiness, General Services; EEI = Energy and Extractive Industries, FINCAP = Finance and Capital Markets; INF = Infrastructure; MIGA = Multilateral Investment Guarantee Agency.

MIGA has enhanced its self-evaluation efforts. MIGA's ability to collect information and track development results is inherently limited by its role as a provider of guarantees. However, MIGA has minimized these challenges by having teams visit nearly all projects that are subject to self-evaluation. MIGA has made efforts to self-evaluate all of its projects, with IEG only doing validations. MIGA has been deferring the evaluations for projects that are not yet fully operational and projects with political risks until those issues are resolved and its criteria for determining which projects are eligible for evaluation have become firmer.[1]

Access to goods and services for customers and market development are prominent outcome types for MIGA. To assess longer-term investment rating trends, we carried out an outcome type analysis, which differs from the analysis we carried out for IFC in that it could not rely on retroactively applied Impact Measurement and Project Assessment Comparison Tool data (box 4.1). MIGA's outcome types can be divided among project-level outcomes and foreign investment–level outcomes. Among outcome types, "access to goods and services for customers" and "market development" are the most common, accounting for 70 and 47 percent of projects (appendix C, table C.8). Both of those outcome types are typical for large infrastructure projects that promote foreign investment. MIGA has an increasing share of foreign investment outcome types, such as "improved business and sector practices" and "signaling effects." These outcome types are expected to have demonstration effects.

MIGA projects have a higher probability of achieving project-level outcomes than foreign investment–level outcomes. As described in chapter 3, the achievement of project-level outcomes tends to be under the direct control of projects. Our analysis shows that the project-level outcome types of "access to goods and services for customers," "quality and affordability of goods and services," and "increased employment" have the highest probabilities of success. From FY12–14 to FY17–19, project-level outcome achievement rates increased (table 4.1). MIGA's inherent limitations as a guarantee provider in collecting data on the development results of projects means that many projects lack sufficient evidence to rate project outcomes. This suggests that MIGA's improved development outcome ratings are due to both increased evidence collection in recent years and actual improvement in performance.

## Box 4.1. MIGA Outcome Type Analysis Methodology

Chapter 4 analyzes the influence certain outcome types have on the performance of Multilateral Investment Guarantee Agency (MIGA) projects.. The outcome typology includes 13 outcome types that were adapted from MIGA's Impact Measurement and Project Assessment Comparison Tool (IMPACT) framework. The analysis reviewed all MIGA projects that were evaluated during fiscal year (FY)12–14 and FY17–19 and validated by the end of calendar year 2020. The *Results and Performance of the World Bank Group 2021* team coded each outcome claim from the Board proposal (President's Report) and assessed outcome claim achievement by reviewing Project Evaluation Reports and Validation Notes. The way claim achievement was assessed was different from the approach of IMPACT and more streamlined. The intention in this chapter was to analyze outcome types in a similar manner to chapters 2 and 3, but this was constrained by the fact that the President's Report included a large number of outcome claims without specific indicators to verify those claims. Moreover, MIGA did not backfill project information with IMPACT data, unlike the International Finance Corporation, so there were certain difficulties in specifying intended outcome claims and seeing the claim rating change. This created a risk that outcome types would not be assigned objectively by this analysis team. Hence, the results of MIGA's outcome type analysis should be interpreted cautiously. Appendixes A and B provide further details on the methodology.

*Source:* Independent Evaluation Group.

Aligning projects with certain corporate priorities had an inconclusive impact on ratings. The *RAP 2020* identified several factors associated with MIGA's relatively higher ratings. These factors include larger multilateral investors, larger project sizes, and beginning MIGA's involvement in projects once the projects are more advanced. We used Project Evaluation Report ratings to analyze the performance of projects that feature corporate priorities. In addition to MIGA's mandate of promoting foreign investment, MIGA's three-year rolling strategies have identified four prominent corporate strategies, including IDA, FCS, climate change, and South-South investments (see box 4.2 for more details on MIGA's corporate priorities). Figure 4.3 shows that the performance of projects with and without a focus on corporate priorities is not obviously different over the past decade. Considering the small number of MIGA projects, the quantitative analysis would need to be complemented by qualitative analysis to shed further light on these findings.

### Table 4.1. Outcome Type Performance, Evaluation Stage, 2012–14 and 2017–19

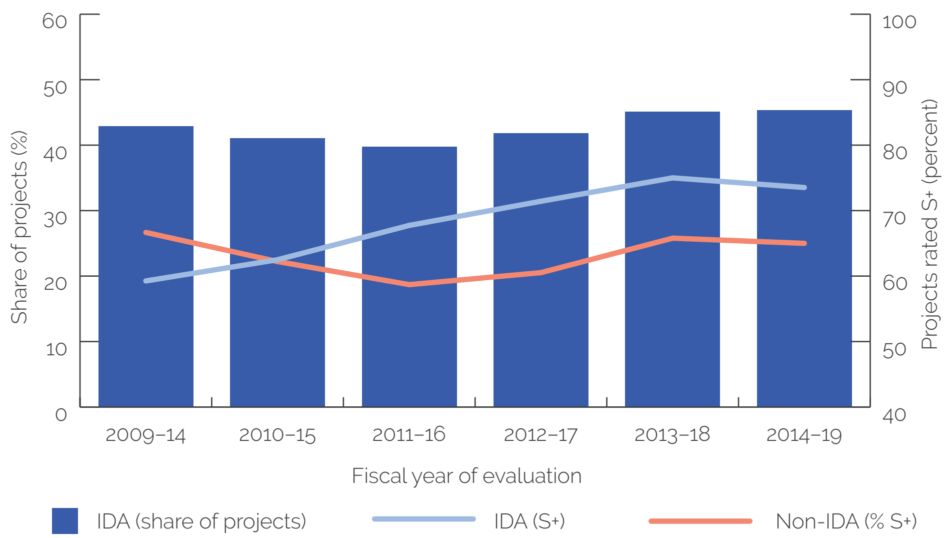
| Outcome Type Performance | 2012–14 Outcome Claims (% achieved) | 2012–14 Outcome Claims (no.) | 2017–19 Outcome Claims (% achieved) | 2017–19 Outcome Claims (no.) |
|---|---|---|---|---|
| Project-level outcome | | | | |
| 1.1 Access (1.1.1–1.1.3) | 33 | 27 | 56 | 54 |
| 1.1.1 Access to goods and services (MSME) | 11 | 9 | 56 | 9 |
| 1.1.2 Access to goods and services (female) | 0 | 2 | — | 0 |
| 1.1.3 Access to goods and services (customers) | 42 | 19 | 57 | 46 |
| 1.2 Quality/affordability of goods and services | 46 | 13 | 52 | 29 |
| 1.3 Enhanced capacity of final beneficiaries | 33 | 9 | 50 | 2 |
| 1.4 Improved living standards (earnings) of individuals | — | 0 | 100 | 2 |
| 2.1 Suppliers/distributors reached | 0 | 2 | 100 | 2 |
| 2.3 Improved sales/profitability of suppliers/distributors | 0 | 5 | 33 | 9 |
| 3.1 Increased employment | 54 | 13 | 38 | 21 |
| 4.1 Increased transfers to the government | 30 | 10 | 33 | 18 |
| 6.2 GHG reduction | — | 0 | 57 | 7 |
| 6.3 Efficient use of resources | 40 | 5 | 71 | 7 |
| Foreign investment–level outcome | | | | |
| 9. Business & sector practices | 40 | 5 | 40 | 15 |
| 10. Market development | 39 | 23 | 29 | 24 |
| 13. Signaling effects | — | 0 | 100 | 1 |

*Source:* Independent Evaluation Group.

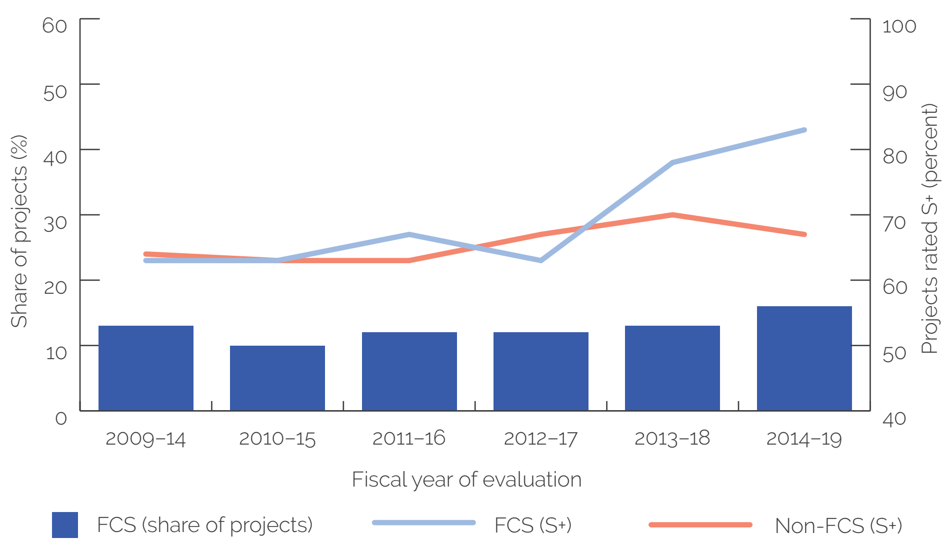
*Note:* Outcome type estimation by Independent Evaluation Group based on the approval documents and confirmed cases at the evaluation. Of the outcome claims considered not achieved, 51 percent were considered not achieved because their results could not be verified. Broken down by outcome types, foreign investment–level outcome claims had higher share of unverified claims (61 percent) than project-level outcome claims (50 percent). The level is relatively high, particularly for foreign investment–level outcome claims, and it would be expected that the tracking of outcome claim results is strengthened under the newly introduced Impact Measurement and Project Assessment Comparison Tool framework. — = not applicable; GHG = greenhouse gas; MSME = micro, small, and medium enterprise.

Figure 4.3. MIGA Projects with Development Outcomes Rated S+, with and without Corporate Priority Objectives

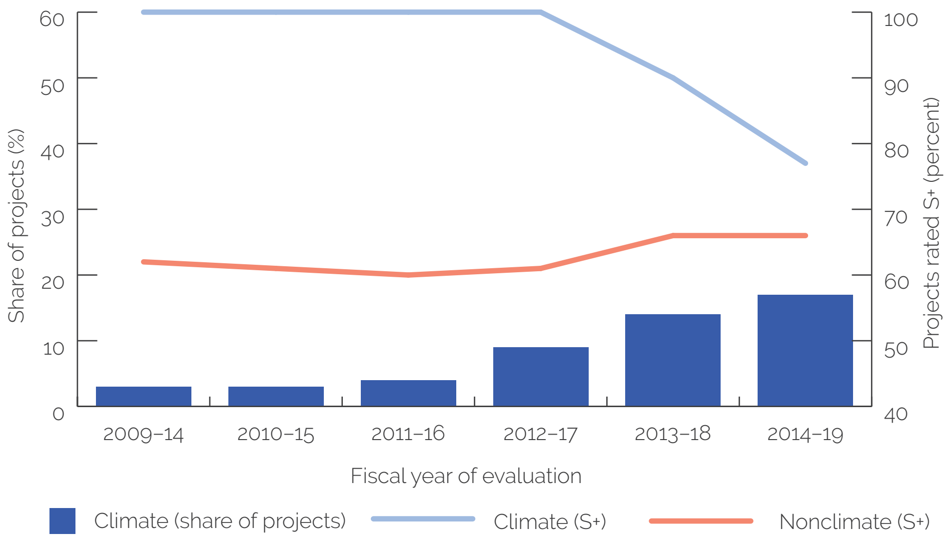
a. Performance of IDA projects



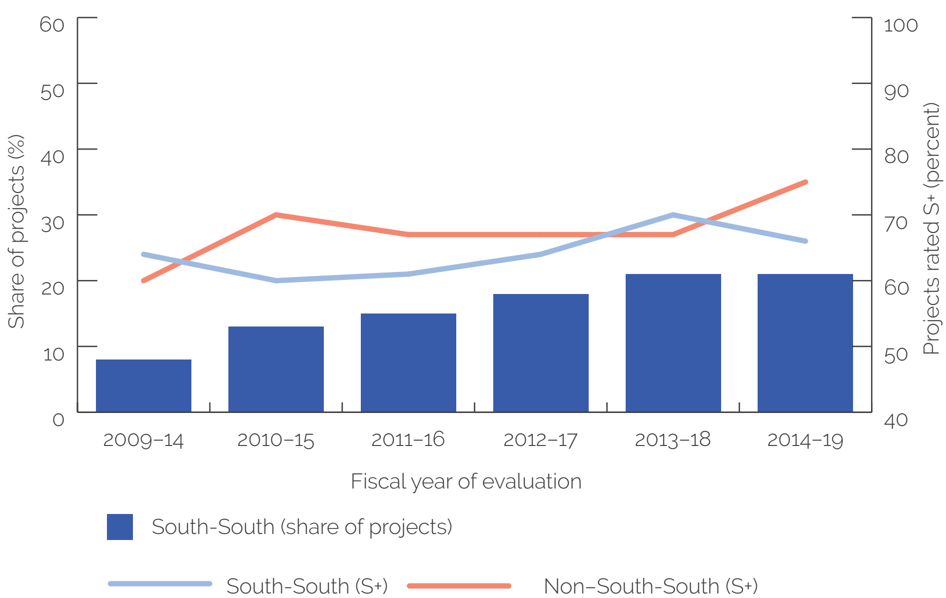

b. Performance of FCS projects



c. Performance of climate change projects




d. Performance of South-South projects




*Source:* Independent Evaluation Group.

*Note:* FCS = fragile and conflict-affected situation; IDA = International Development Association; MIGA = Multilateral Investment Guarantee Agency; S+ = satisfactory or above.

## Box 4.2. MIGA's Select Corporate Priorities

**International Development Association countries.** The Multilateral Investment Guarantee Agency (MIGA) introduced International Development Association (IDA) countries as a strategic priority area for in 2005; this has continued to be a priority area ever since.

**Fragile and conflict-affected situations.** Fragile and conflict-affected situation (FCS) countries have been a strategic priority for MIGA since 2005, and MIGA's current strategy reinforces this focus. MIGA's fiscal year (FY)21–23 strategy aims to increase the share of MIGA guarantees in IDA and FCS countries to an average of 30 to 33 percent during FY21–23.

**Climate change.** The World Bank Group launched its first Climate Change Action Plan in 2016 and has published a second plan for 2021–25. In the second plan, MIGA, as part of the Bank Group, is committed to aligning all new operations with Paris Agreement goals by 2025.

**South-South investment.** MIGA supports South-South investments by promoting foreign investments from developing countries. MIGA made this a strategic focus area in its FY09–11 and FY12–14 strategies (the FY15–17 strategy mentioned it as an area of support).

**Methodology.** To assess the relationship between projects that integrate corporate priorities and their Project Evaluation Report ratings, we classified all the projects evaluated since the FY09 cycle using climate change flags provided by MIGA, South-South flags provided by the Independent Evaluation Group, and FCS classification at the time of evaluation.[a] A project can have multiple corporate priorities.

*Source:* Independent Evaluation Group.

*Note:* MIGA = Multilateral Investment Guarantee Agency.
a. Subject to the availability of data at the time of assessment. For country classification for IDA and FCS, historical country classification based on the Bank Group's classification was applied for evaluation year of projects. For example, for a project in the 2016 evaluation year, IDA and FCS country classifications of the Bank Group in 2016 were applied.

[1] The point in time at which the criteria for evaluation have been met is called "early operating maturity."

# 5. Conclusions

The recent increases in the World Bank's project outcome ratings and IFC's development outcome ratings are positive news. The World Bank's outcome ratings steadily improved from FY10 onward before increasing by an impressive 9 percentage points in FY20, reaching 88 percent of projects with outcome ratings of MS or higher, a historic high. The increase resulted from ratings improvements for virtually all categories of projects—all Practice Groups (especially Sustainable Development), all Regions (especially Europe and Central Asia and Western and Central Africa), and almost all lending sizes (especially the largest projects of $100 million or more)—rather than resulting from shifts in portfolio composition or improvements limited to specific portfolio segments. Ratings even increased in IDA and FCS countries, the most difficult operating environments. Our analysis shows that disruptions caused by COVID-19 did not have a discernable impact on the ratings jump during FY20. IFC and MIGA saw ratings improvements as well. In 2019, IFC's ratings increased for the first time in 10 years, though there is not enough data to confirm if this improvement was sustained in 2020. MIGA's project development outcome ratings have been steadily increasing for 10 years. Ratings increases across the Bank Group signal the institutions' ongoing commitment to development effectiveness. What the analysis in this report also shows is that project ratings alone provide little evidence on the types of outcomes the Bank Group is achieving and the quality of associated targets and indicators.

The analysis in this report shows that although the presence and implementation of project-level M&E frameworks has improved, many World Bank projects still do not adequately measure the outcomes. Targets and indicators are a critical element of the World Bank's self-evaluation methodology and are a key driver of the assessment of how well projects perform. But the logic and quality of these targets and indicators varies widely, and the correlation between ratings and the quality of targets and indicators is inconsistent, meaning that projects can still achieve high efficacy ratings even when they lack proper baselines or when they measure outputs and activities that do not match a project's intended outcomes.

» Implication: The World Bank and IEG could pay more attention to how well indicators measure project objectives. To do so would require a more systematic approach to gauging the appropriateness of indicators and targets early in the project cycle. A successful approach would include clarifying the links between indicators and project objectives and defining targets in relation to scrutinized baselines. The World Bank's recent ICR reforms, which require an explicit reference to theories of change, are a step in the right direction. The World Bank could, however, make further efforts to select robust, direct, and attributable indicators and targets.

Like any other metric, aggregate ratings need to be interpreted correctly by understanding what they do and do not measure. As discussed above, ratings measure a project's success in meeting self-defined targets and objectives, but ratings are not meant to assess either the nature of a project's development outcomes or the extent to which the project addresses a country's development needs. This means that individual ratings use indicators at different levels of ambition and complexity and are not measured by an absolute standard. As Bulman, Kolkma, and Kraay (2015) observe, this introduces the possibility that at least some of the variation in aggregate project outcome ratings is caused by differences in the ambition or attainability of the stated development objective, rather than any differences in actual outcomes. This problem is less acute in IFC and MIGA because their evaluation framework includes some objective criteria and standards (such as a project's financial performance and comparisons with peers and industry benchmarks).

» Implication: The World Bank could provide a fuller explanation of ratings as and how they relate to underlying development outcomes. IEG and the World Bank could carry out periodic syntheses and report on development outcomes, following in the footsteps of IEG's outcome orientation agenda. Potentially, the World Bank could devise a system to regularly harvest project outcomes and key activities and match this information with ratings data for a more integrated monitoring of results and performance.

A project's development outcomes are affected by a host of factors not directly considered in ratings. Ratings serve a strong purpose in evaluating a project's performance, but supplementing ratings with information about a project's size, type, country, outcome type, client type, outcome potential, corporate priorities, and other characteristics can help teams attain a fuller, more objective assessment of a project's development outcome and its risk—as IFC and MIGA analysis has shown. This report shows that some of these characteristics may have a direct impact on ratings (for example, IFC projects with repeat clients tend to have higher ratings), other characteristics have a tenuous link to ratings (for example, pursuing certain corporate priorities), and still other characteristics have no confirmed effect on ratings (for example, outcome types on overall ratings). This is not to say, however, that these characteristics do not provide context to ratings and help project teams better understand a project's probability of success or the risk for achieving certain development outcomes.

» Implication: IFC and MIGA could use information on outcome types and other characteristics to better assess projects' risks, ratings, and development outcomes. IFC's AIMM framework and MIGA's Impact Measurement and Project Assessment Comparison Tool framework already account for a project's estimated and actual development potential and development outcome risks. IFC and MIGA could take it a step further by assessing the prevalence of different outcome types and other characteristics in projects to help enhance their frameworks. For example, the potential risk severity of outcome types, as manifested in ratings variance—and the difficulty of achieving certain outcome types—can be incorporated into a project's development outcome and risk assessment. Adding this information to a typical assessment of development outcomes would contribute to the Bank Group's learning and possibly improve the ratings system itself.

High ratings do not appear to signal risk aversion. Early in the *RAP* process, we hypothesized that ratings increases could have resulted from operational teams taking less risk. Indeed, the *RAP 2020* and past evaluations identified this as a potential danger (World Bank 2016a, 2020c). However, this report shows that in the two GPs analyzed—Transport and Education—successor projects that introduced novelty (that is, that introduced new or expanded elements over the previous project) performed as well as or better than projects that closely replicated the predecessor project. The analyses of IFC's XPSR ratings of projects with market-level outcomes and the relationship between outcome potential and XPSR ratings suggest that projects addressing high-magnitude development outcomes are not destined for lower ratings and that a project's outcome potential and XPSR rating may actually be moving in the same direction. All this suggests that the World Bank and IFC operational teams can discern when the conditions are right for projects to support novel and complex activities. In this way, teams can take informed risks and selectively build on past experiences to elevate a project's objectives without suffering lower project performance ratings.

» Implication: The Bank Group could further emphasize operating "on the frontier" (that is, selecting the best combination of risks and opportunities) as a goal in addition to meeting the Corporate Scorecards rating targets. This shift in emphasis would provide a broad set of incentives and encourage the Bank Group to inquire further about the motivations for risk taking; the evolution of project designs; the pursuit of corporate priority goals; and the best way to leverage internal resources and the client's engagement, commitment, and capacity to deliver development results. This could help ensure that the Bank Group continues to selectively take risks to improve development outcomes.

# Bibliography

Bulman, David, Walter Kolkma, and Aart Kraay. 2015. "Good Countries or Good Projects? Comparing Macro and Micro Correlates of World Bank and Asian Development Bank Project Performance." Policy Research Working Paper 7245, World Bank, Washington, DC.

Chauvet, Lisa, Paul Collier, and Marguerite Duponchel. 2010. "What Explains Aid Project Success in Post-Conflict Situations?" Policy Research Working Paper 5418, World Bank, Washington, DC.

Legovini, Arianna, Vincenzo Di Maro, and Caio Piza. 2015. "Impact Evaluation Helps Deliver Development Projects." Policy Research Working Paper 7157, 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.

Smets, Lodewijk, Stephen Knack, and Nadia Molenaers. 2013. "Political Ideology, Quality at Entry and the Success of Economic Reform Programs." Policy Research Working Paper 6130, World Bank, Washington, DC.

World Bank. 2008. "Ghana—Road Sector Development Project." Implementation Completion and Results Report ICR806, World Bank, Washington, DC. http://documents.worldbank.org/curated/en/361401468250510275/Ghana-Road-Sector-Development-Project.

World Bank. 2012. "Argentina—Additional Financing for the Buenos Aires Urban Transport Project." Implementation Completion and Results Report ICR2357, World Bank, Washington, DC. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/628281468004776791/argentina-additional-financing-for-the-buenos-aires-urban-transport-project

World Bank. 2015. *Results and Performance of the World Bank Group 2014*. Independent Evaluation Group. Washington, DC: World Bank. https://ieg.worldbankgroup.org/evaluations/rap2014.

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 2016*. Independent Evaluation Group. Washington, DC: World Bank. https://ieg.worldbankgroup.org/evaluations/results-and-performance-2015.

World Bank. 2017a. "Nepal—School Sector Reform Program Project." Implementation Completion and Results Report ICR3994, World Bank, Washington, DC. http://documents.worldbank.org/curated/en/794561490279193590/Nepal-School-Sector-Reform-Program-Project.

World Bank. 2017b. "Nepal—Vocational Education Project." Implementation Completion and Results Report ICR4135, World Bank, Washington, DC. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/458241514911318490/nepal-vocational-education-project.

World Bank. 2018a. *The International Finance Corporation's Approach to Engaging Clients for Increased Development Impact*. Independent Evaluation Group. Washington, DC: World Bank. https://ieg.worldbankgroup.org/evaluations/ifc-client-engagement.

World Bank. 2018b. *Results and Performance of the World Bank Group 2017*. Washington, DC: World Bank. https://ieg.worldbankgroup.org/evaluations/rap2017.

World Bank. 2019a. *Creating Markets to Leverage the Private Sector for Sustainable Development and Growth*. Independent Evaluation Group. Washington, DC: World Bank. https://ieg.worldbankgroup.org/evaluations/creating-markets.

World Bank. 2019b. *Ghana—Transport Sector Project*. Washington, DC: World Bank Group.

World Bank. 2019c. *Results and Performance of the World Bank Group 2018*. Washington, DC: World Bank. https://ieg.worldbankgroup.org/evaluations/rap2018.

World Bank. 2020a. *Argentina—Urban Transport in Metropolitan Areas Project*. Washington, DC: World Bank Group.

World Bank. 2020b. "Operations Policy and Country Services' Bank Guidance: Implementation Completion and Results Report for Investment Project Financing (IPF) Operations." March 2, 2020.

World Bank. 2020c. *Results and Performance of the World Bank Group 2020*. Independent Evaluation Group. Washington, DC: World Bank. https://ieg.worldbankgroup.org/evaluations/rap2020.

World Bank. 2020d. *The World Bank Group Outcome Orientation at the Country Level*. Independent Evaluation Group. World Bank: Washington, DC.

World Bank. 2021a. *How to Improve Results in FCV Environments: 12 Recommendations*. Washington, DC: World Bank.

World Bank. 2021b. *Results and Performance of the World Bank Group 2021*. Concept Note. Independent Evaluation Group. Washington, DC: World Bank.

World Bank. Forthcoming. *IFC's and MIGA's Support for Private Investment in Fragile and Conflict-Affected Situations (FY10–21)*. Independent Evaluation Group. Washington, DC: World Bank.

# APPENDIXES

*Results and Performance of the World Bank Group 2021*

# Appendix A. Methodological Approach

The report's overarching question is, What does the existing evidence from the Independent Evaluation Group's (IEG) project evaluation and validation work show about the World Bank Group's results and performance?

To answer this question, the *Results and Performance of the World Bank Group 2021 (RAP 2021)* report developed several blocks of analysis, as follows:

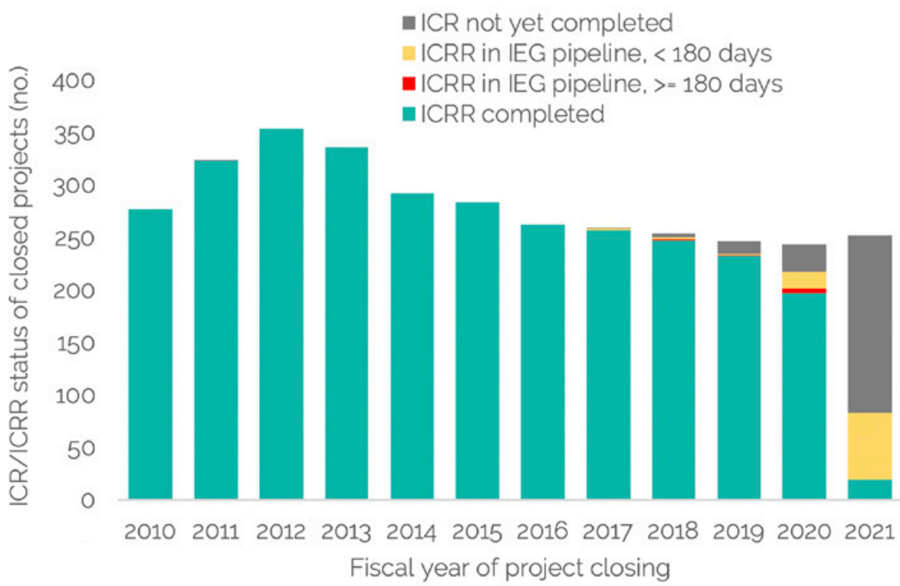
- Analysis of ratings (World Bank, International Finance Corporation [IFC] investments, IFC advisory services, Multilateral Investment Guarantee Agency [MIGA]);
- Analysis of outcomes (World Bank, IFC investments, MIGA); and
- In-depth select analysis (for introduction of novelty in projects, indicators, and targets).

## Analysis of Ratings

### World Bank

The overall rating trend includes all projects that closed between fiscal years (FY)00 and FY20 and had an Implementation Completion and Results Report (ICR) and ICR Review (ICRR) completed by August 10, 2021 (*N* = 5,825), with a special focus on the most recent period, 2010–20 (*N* = 3,080). As of that date, the FY20 cohort includes 194 projects, which is 91 percent of the ICRs that IEG has received for projects that closed in FY20, or 81 percent of the total including ICRs expected that have not yet been completed (figure A.1). For projects that closed in FY17 or later, the coverage is 97 percent of ICRs received by IEG, and 93 percent including ICRs not yet completed for closed projects.

**Figure A.1.** Coverage of World Bank Project Ratings Data for This Report




*Source:* Independent Evaluation Group.

*Note:* ICR = Implementation Completion and Results Report; ICRR = Implementation Completion and Results Report Review; IEG = Independent Evaluation Group.

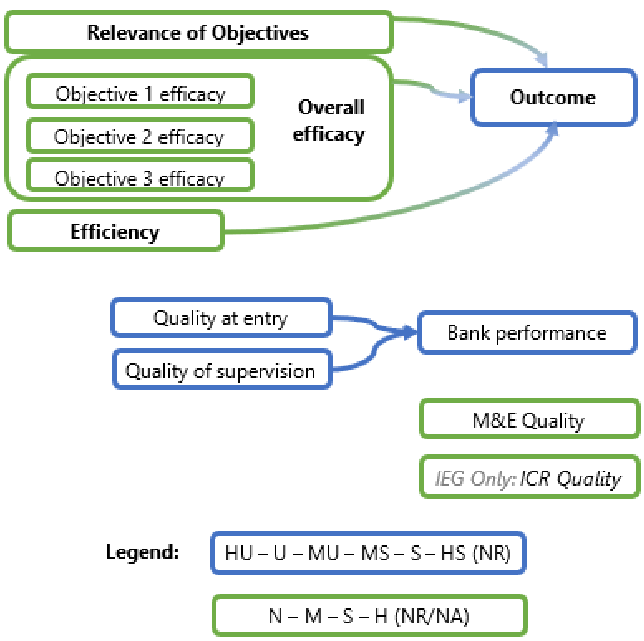
The ratings used to indicate World Bank project performance in this report are based on IEG's validation reviews of ICRs (that is, ICRRs) completed by operational teams after projects close.

An ICR is prepared by the World Bank at the close of every operation funded by the International Development Association or the International Bank for Reconstruction and Development or, in the case of a series of programmatic policy operations, at the end of that series. An ICR is expected to constitute a complete and systematic account of the performance and results of the project. In addition to telling the project's results story, the ICR contains ratings of the project's performance. The ratings often used in monitoring are the outcome rating, which is based on the subratings of the project's relevance, efficacy, and efficiency, and the Bank performance rating, which takes into account the World Bank's performance in ensuring quality at entry and its performance in supervision of the project. Ratings scales and criteria were developed through collaboration between the World Bank's Operations Policy and Country Services and IEG and have evolved somewhat over time.

The ICRR conducted by IEG is an independent, desk-based, critical review of the evidence, results, and ratings of the ICR in relation to the project's design documents. Based on the evidence provided in the ICR and an interview with the final task team leader, IEG arrives at its own ratings for the project based on the same evaluation criteria used by the World Bank (figure A.2).

In reviewing the findings and ratings in the ICR, IEG provides an independent view of the results and ratings, conditioned on both the evidence presented in the ICR and the evidence provided by the final task team leader for the project. However, IEG is not privy to evidence that is not included in the ICR. The ICRR is thus an independent validation of the World Bank's self-evaluation and ratings; it is not an independent evaluation of the project based on evidence collected outside the World Bank's self-evaluation process.

Figure A.2. Ratings Elements in the Independent Evaluation Group's Implementation Completion and Results Report Review—Investment Project Financing Example




*Source:* Independent Evaluation Group.

*Note:* This is an example for an IPF with no restructuring and a project development objective that contains three individual objectives. Elements outlined in blue are rated on a six-point scale (from lowest to highest: highly unsatisfactory [HU], unsatisfactory [U], moderately unsatisfactory [MU], moderately satisfactory [MS], satisfactory [S], and highly satisfactory [HS]). Elements outlined in green are rated on a four-point scale (from lowest to highest: negligible [N], modest [M], substantial [S], high [H]). ICR = Implementation Completion and Results Report; IEG = Independent Evaluation Group; IPF = investment project financing; M&E = monitoring and evaluation.

# International Finance Corporation

## International Finance Corporation Investment Project

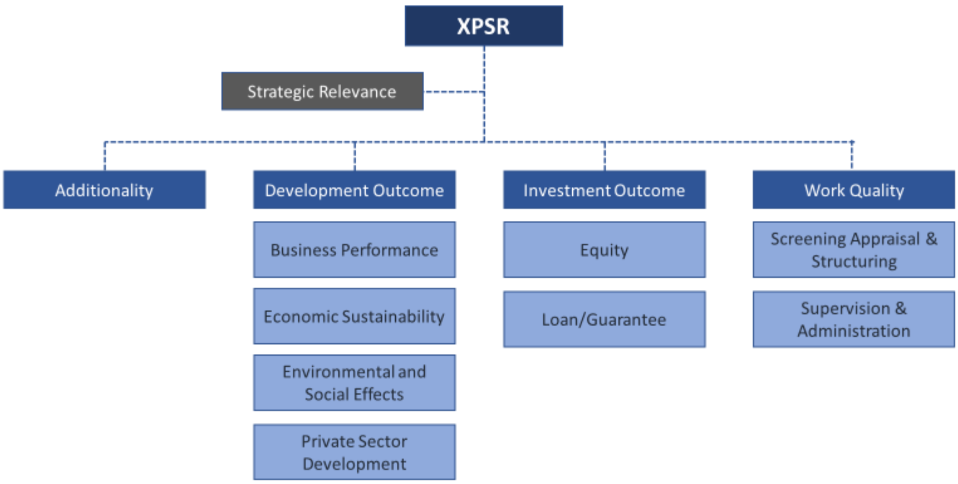
IFC's project evaluation system is a self-evaluation system based on the Expanded Project Supervision Report (XPSR), independently validated by IEG. The XPSR system enables IFC (i) to be accountable to its Board and shareholders in terms of its purpose, which is to further economic development by encouraging the growth of productive private enterprises in member countries, thus supplementing the activities of the World Bank; (ii) to contribute to learning through the identification of lessons and updated information on conditions and prospects to improve current operations and strategy; and (iii) to provide independently validated development results that feed into metrics for corporate, department, and individual performance.

Selection. XPSRs are prepared for a representative random sample of 40 percent of mature operations, selected by IEG each year and announced in December. For this *RAP*, the analysis of the latest period is based on a stratified random representative sample, which for calendar years (CY)17–19 covered 270 projects, or 40 percent of all projects approved in CY12–14.

IFC's investment staff complete the self-evaluation based on the joint IFC and IEG guidelines that define the content and scope of XPSRs and the criteria for attributing indicator and outcome ratings. IEG's Financial and Private Sector Micro Unit validates the XPSR findings and lessons and independently rates the development results summarized in IEG's Evaluation Note (EvNote).

Evaluation framework. An XPSR contains a rating of the project's emerging development results (based on business performance, economic sustainability, environmental and social effects, and private sector development), IFC's investment performance, IFC's operational effectiveness (work quality), and additionality (figure A.3). It assesses the project's strategic relevance and the achievement of the project's stated objectives presented in the Board report at approval. It compares the project's performance with relevant competitors or sector benchmarks for the specific industry or niche. It also includes an analysis of the prospects of the operation to assess the sustainability of the results in the longer term.

Figure A.3. Performance Areas and Dimensions of Expanded Project Supervision Reports




*Source:* Independent Evaluation Group.

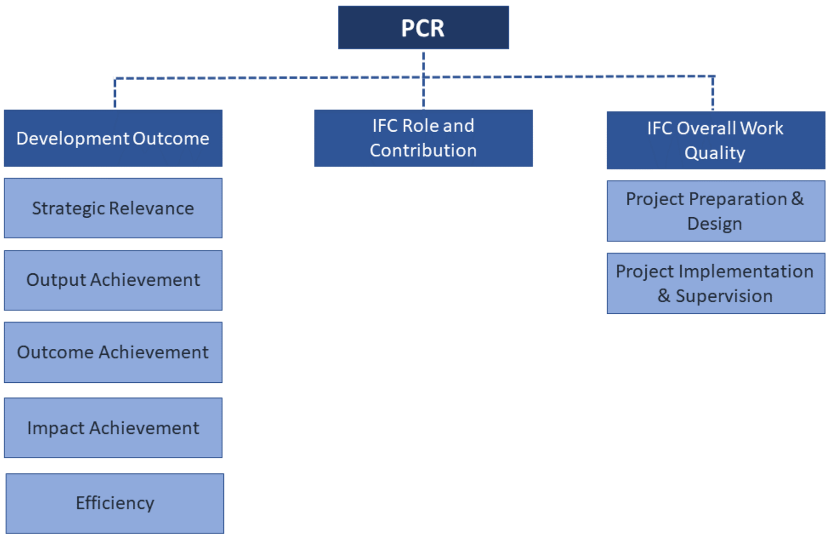
*Note:* XPSR = Expanded Project Supervision Report.

## International Finance Corporation Advisory Project

IEG's independent evaluation encompasses independent validation of advisory services Project Completion Reports, which first launched in FY08. Validation is a similar level of review to that which IEG carries out for investment operations under IFC's XPSR system (figure A.4). In addition to assessing the evaluative integrity of the Project Completion Reports, IEG independently validates the report's findings and ratings. It does this through an in-depth desk review of all project documentation and involves greater access to project files and external or independent sources of information. It also involves discussions with project teams and, as necessary, clients and other relevant stakeholders (by phone). Where appropriate, IEG supplements the desk-based validation with a field visit to observe the results in the field and interview the project's clients and stakeholders in person. Independent validation focuses on a sample of projects rather than the whole population. IEG does carry out an internal peer review to ensure inter-rater consistency.

For this *RAP*, the analysis of the latest period is based on a random representative sample, which for FY17–19 covered 184 projects, or 50 percent of all projects completed in FY17–19.

Figure A.4. Performance Areas and Dimensions of Project Completion Reports




*Sources:* Independent Evaluation Group and International Finance Corporation.

*Note:* IFC = International Finance Corporation; PCR = Project Completion Report.

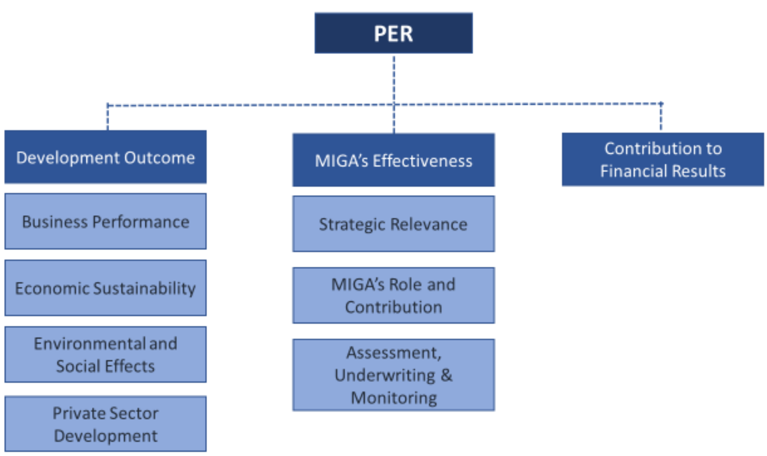
## Multilateral Investment Guarantee Agency

MIGA's project evaluation system is a self-evaluation system based on the Project Evaluation Report (PER), which is independently validated by IEG. The PER system enables MIGA (i) to be accountable to its Board and shareholders regarding its purpose, which is to promote foreign direct investment into developing countries to help support economic growth, reduce poverty, and improve people's lives, thus supplementing the activities of the World Bank and IFC; (ii) to contribute to learning through the identification of lessons and updated information on conditions and prospects to improve current operations and strategy; and (iii) to provide independently validated development results that feed into metrics for corporate, department, and individual performance.

Selection. All eligible, mature MIGA projects are subject to self-evaluation via the PER system. In the past few years, the annual PER program has covered between 10 and 20 projects. For active operations at the time of evaluation, MIGA's staff complete the self-evaluation based on IEG-MIGA guidelines that define the content and scope of PERs and the criteria for attributing indicator and outcome ratings. IEG's Financial and Private Micro Sector Unit validates the PER findings and lessons and independently rates the development results summarized in IEG's EvNote.

Evaluation framework. A PER contains a rating of the project's emerging development results (based on business performance, economic sustainability, environmental and social effects, and private sector development), MIGA's effectiveness (strategic relevance; role and contribution; and assement, underwriting, and monitoring), and the project's contribution to MIGA's financial results (figure A.5). It assesses the project's strategic relevance and achievement of the project's stated objectives presented in the Board report at approval. It compares the project's performance with relevant competitors or sector benchmarks for the specific industry or niche. It also includes assessment of the intended objectives and analysis of the prospects of the operation to determine the sustainability of the results in the longer term. The MIGA PER guidelines were updated in 2019, and all the project evaluations reported in this *RAP* were prepared under the previous guidelines.

Figure A.5. Performance Areas and Dimensions of Project Evaluation Reports




*Source:* Independent Evaluation Group.

*Note:* MIGA = Multilateral Investment Guarantee Agency; PER = Project Evaluation Report.

For this year's *RAP*, all projects that were eligible for evaluation were evaluated. The analysis of the latest year included 75 projects that were evaluated during the six-year period (FY14–19).

## Outcome Types

### World Bank

The outcome types were defined to capture the type of change envisioned by project objectives. Each individual project objective was used as a unit of analysis as identified at the time of validation (at the time of validation, the ICRR validator parses the project development objective into individual objectives).[1]

Sixteen outcome types were identified for the World Bank, based on the following criteria:

» Analysis of the typical theories of change of World Bank projects to define common outcomes that projects are set up to achieve. This step enabled identification of key categories such as expanded access to services; improved service quality; increased human capital; improved individual employability and livelihood; improved enterprise and sectoral performance; and enhanced accountability, transparency, or governance.

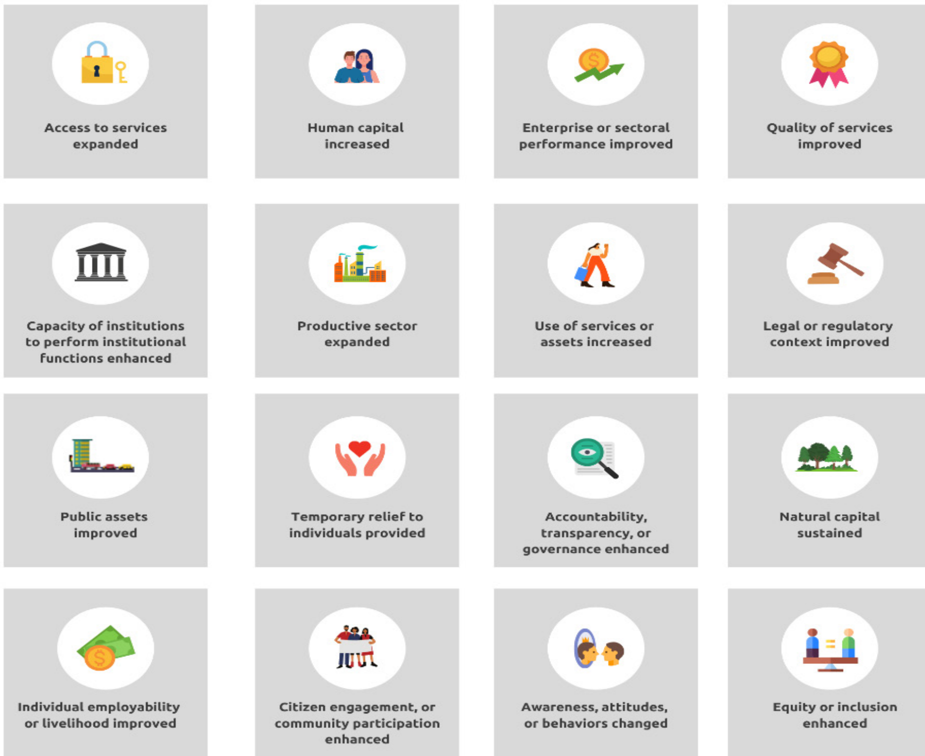
» Identification of additional categories to ensure that all project objectives included in the sample were mapped to the appropriate theory of change and outcome types. This step enabled the definition of a few categories initially overlooked (such as improved public assets).

» Identification of some general outcome types reflected in corporate goals (such as increased human capital and enhanced equity and inclusion).

» Identification of qualitative markers of project objectives, which do not typically represent the main outcome identified by the explicitly stated objective, but capture salient features of the project emerging from the analysis of project activities or indicators. Examples of these categories are increased equity and inclusion, enhanced citizens' engagement or community participation, and changed awareness, attitudes, or behaviors. Occasionally, these outcome types can be the main and only outcome type.

One individual project objective could be classified into several outcome types. Also, outcome types were conceptualized as not mutually exclusive (figure A.6).

Figure A.6. The World Bank's 16 Outcome Types




*Source:* Independent Evaluation Group.

The approach followed to decide which outcome type(s) to assign to each objective was to consider, in order of priority, the *objective statement*, the (outcome and output) *indicators*, and the *activities*. Solely relying on the literal objective statement was deemed insufficient and potentially misleading—the same formulation of an objective statement in two different projects can represent very different intended outcomes; conversely, two different formulations could indicate the same intended outcomes. At the same time, just the indicators could not determine the outcome type, as indicators may be inappropriately selected to measure the results of an objective (as is clear from the low IEG monitoring and evaluation ratings). That said, indicators often contributed to making or confirming a decision about outcome types, especially when dealing with convoluted objective statements.

The strengths of the approach chosen and the difficulties and the limitations encountered in the classification of outcome types are noted in box A.1.

### Box A.1. Strengths and Limitations in Identifying Objective Types

#### Strengths

Cross-cutting nature of outcome types: The outcome types represent an explicit description of the intended changes of the objectives, based on the theory of change inherent to the project. They cut across Practice Groups and Global Practices and could allow for sharing evidence and learning across projects aiming at similar types of outcomes, even when the sector is different.

Granularity: Each individual objective present within the project development objective was coded to account for all the changes pursued by the project.

Changes over time: Objectives were coded both at entry (start of the project) and at exit (project closing) to capture changes in objectives that may have occurred at restructuring and allow for correct matching of each objective to the corresponding efficacy rating.

Multiple coding: Multiple coding of the same objective was allowed so that coding is minimally dependent on the typology adopted (for example, it is possible to introduce new outcome types without affecting the existing coding).

#### Difficulties or limitations encountered with some objectives

Unclear theory of change: This limitation is especially apparent in older projects, which privilege the description and justification of activities over the explanation of the intervention logic based on a plausible theory of change that links activities to objectives.

Complex objectives, which often point to multiple goals: The decision on how to parse the project development objective into its separate components is made by the Independent Evaluation Group evaluator, whose skills and expertise determine how logical and well defined the individual project objectives are.

*(continued)*

### Box A.1. Strengths and Limitations in Identifying Objective Types (cont.)

Unrealistic objectives: These describe a goal that cannot be plausibly achieved by or attributed to the project (such as decreasing poverty in a country).

Limited information in the Implementation Completion and Results Report Review (the document used to identify outcome types): These reviews do not match project outcome indicators with specific objectives, which would help with assigning outcome types to objectives. In addition, if no project activities have taken place (which may happen even in the absence of formal restructuring), the review does not generally report on intended outcomes. Details on project activities vary. Inconsistencies exist in the quality of validation.

Lack of adequate indicators: This makes it harder to discern outcome types, especially when the objective is poorly stated.

Disconnect between indicators and activities or objectives: This introduces ambiguities in the definition of outcome types.

*Source:* Independent Evaluation Group.

The analysis of outcome types was carried out for a representative sample of World Bank projects for two different time periods (projects that closed in FY14–16 formed the earlier sample and projects that closed from FY17 to the second quarter of FY20 formed the later sample).[2] The sample was drawn from data available as of January 21, 2021. The decision to sample was taken because of the limited resources available. The two samples were representative at the Global Practice (GP) family level (90 percent level, 10 percent margin of error). Representative samples were also created for the specific Education and Transport GPs to conduct more in-depth analysis of both outcome types and novelty (see chapter 2). The sample composition is shown in table A.1.

**Table A.1.** Sample Selection for Analysis of Outcome Types

| Global Practice or Practice Group | Population (N = 1,490) | | Sample (N = 448) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | FY12–14 | | | | FY17–20 (Q1 and Q2) | | | |
| | FY12–14 | FY17–20 (Q1 and Q2) | All | GP Group | GP | Additional | All | GP Group | GP | Additional |
| EFI | | | | | | | | | | |
| Finance, Competitiveness, and Innovation | 47 | 42 | 19 | 19 | | | 19 | 19 | | |
| Governance | 55 | 40 | 17 | 17 | | | 17 | 17 | | |
| Macroeconomics, Trade, and Investment | 16 | 7 | 6 | 6 | | | 1 | 1 | | |
| Poverty and Equity | 3 | 6 | 2 | 2 | | | 3 | 3 | | |
| EFI total | 121 | 95 | 44 | 44 | | | 40 | 40 | | |
| HD | | | | | | | | | | |
| Education | 87 | 72 | 41 | 27 | 39 | 2 | 38 | 23 | 35 | 7 |
| Health, Nutrition, and Population | 74 | 43 | 17 | 17 | | | 18 | 18 | | |
| Social Protection & Jobs | 33 | 32 | 7 | 7 | | | 6 | 6 | | |
| HD total | 194 | 147 | 65 | 51 | | | 62 | 47 | | |

*(continued)*

| Global Practice or Practice Group | Population (N = 1,490) | | Sample (N = 448) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | FY12–14 | | | | FY17–20 (Q1 and Q2) | | | |
| | FY12–14 | FY17–20 (Q1 and Q2) | All | GP Group | GP | Additional | All | GP Group | GP | Additional |
| INFRA | | | | | | | | | | |
| Digital Development | 5 | 10 | 4 | 4 | | | 5 | 4 | | |
| Energy and Extractives | 83 | 75 | 22 | 22 | | | 19 | 19 | | |
| Transport | 78 | 73 | 38 | 23 | 37 | 2 | 37 | 25 | 35 | 3 |
| INFRA total | 166 | 158 | 64 | 49 | | | 61 | 48 | | |
| SD | | | | | | | | | | |
| Agriculture and Food | 67 | 80 | 13 | 13 | | | 15 | 15 | | |
| Environment, Natural Resources, and Blue Economy | 59 | 75 | 11 | 11 | | | 9 | 9 | | |
| Social Sustainabilty & Inclusion | | 4 | | | | | | | | |
| Urban, Resilience, and Land | 111 | 74 | 18 | 18 | | | 17 | 17 | | |
| Water | 66 | 73 | 14 | 14 | | | 15 | 15 | | |
| SD total | 303 | 306 | 56 | 56 | | | 56 | 56 | | |
| Total | 784 | 706 | 229 | 200 | 76 | 4 | 219 | 191 | 70 | 10 |

*Source:* Independent Evaluation Group.

*Note:* EFI = Equitable Growth, Finance, and Institutions; FY = fiscal year; GP = Global Practice; HD = Human Development; INFRA = Infrastructure; SD = Sustainable Development. The two samples for Education and Transport projects were representative at the Global Practice family level (90 percent level, 10 percent margin of error) and extra projects were sampled at the GP level for this purpose. For these two GPs, all projects that were rated highly satisfactory and highly unsatisfactory were also added (column "Additional").

# Methodology of Defining Outcome Types for the International Finance Corporation and Multilateral Investment Guarantee Agency

## International Finance Corporation

### Defining Outcome Types for the International Finance Corporation

For IFC, a 13-category typology of intended outcomes was developed for the *RAP 2021* report, taking advantage of IFC's Anticipated Impact Measurement and Monitoring (AIMM) system, an ex ante project impact assessment tool launched in 2017. The rationale for using the AIMM system for the outcome type analysis is discussed in the following section. However, it should be remembered that the projects to which the AIMM framework was applied from the outset in 2017 have not been validated yet. The main purpose of the use of the AIMM framework was to facilitate the theoretical analysis, which should mirror the approach taken to World Bank outcome analysis.

### Rationale of Using the Anticipated Impact Measurement and Monitoring Framework

Identification of clear outcome claims and their expected results. In addition to the challenges of identifying outcome types explained above, the analysis of IFC projects involved additional challenges because IFC project proposals did not clearly specify intended development objectives before introduction of the AIMM framework. Before the AIMM framework, IFC's project proposal to the Board typically included development indicators, targets and their achievement timeline to highlight to the Board the project's development objectives, and types of development outcomes that would be achieved. The project proposals also tended to include various development outcomes in relatively vague format on multiple levels. In addition, IFC's project proposals included a list of standard indicators, many of which did not measure the development objectives of the project, but a variety of development outcomes. However, when the AIMM system was backfilled to those projects that did not have the AIMM system at the time of approval, IFC's project team identified a limited number of key development outcomes (defined as outcome claims) with specific indicators, in accordance with the theories of changes defined by the AIMM sector framework. Therefore, the AIMM system facilitated the development of an outcome typology and key development claims with relevant indicators for each project.

Provision of ratings relevant to outcome claims. Unlike the World Bank's ICR and ICRR, IFC's self-evaluation tool, XPSR, does not give a performance rating to each project outcome claim. Instead, ratings are provided at four dimensions under development outcome and overall development outcome level. This lack of multiple-scale ratings for outcome claims made the analysis of the riskiness of specific outcome types difficult. However, under the AIMM system, specific ratings (that is, intensity rating for project outcome claim and movement rating for market outcome claim ratings) are provided for each outcome claim, and by comparing the ex ante rating and rating at the monitoring stage, we were able to tell whether actual outcome results exceeded or underperformed the expected results and by how much. In this regard, claim rating difference is more analogous to the ratings that would indicate achievement of outcome objectives. This helped us assess the variance of ratings for specific outcome types.

Assignment of value to development outcome. AIMM ratings or scores are provided based on the country's development needs (gap for the project outcome and market typology for the market outcome) and the extent to which the project contributes to reducing such needs (intensity for the project outcome and movement for the market outcome). As such, the AIMM scores or ratings indicate the level of development challenge that the project is going to address. These ratings are provided not only at outcome level but also at dimension level and overall project level. This allowed us to carry out additional analysis at project dimension and overall project level and assess whether projects to address higher levels of development challenge had lower performance or more variance.

Outcome typology for IFC was developed based on the AIMM sector frameworks. An AIMM sector framework has been prepared for more than 20 key sectors or subsectors of IFC's investment operations. Each sector framework presents an expected theory of change for relevant projects (called the impact thesis), which indicates how the projects in relevant sectors or subsectors are expected to address development gaps. This is done by demonstrating typical outcomes to be achieved by the projects, both at project and market level, and typical types of market gaps relevant to subsectors. The sector framework also includes a list of standard indicators and categorizes them under specific types of outcome. Based on the impact thesis and list of indicators, an outcome typology was developed for 13 outcome categories, some of which had subcategories. IEG added a few categories, such as governance, that were not explicitly included in the sector frameworks.

Outcome claims for each project were identified based on the AIMM backfill worksheet. For those projects with backfilled AIMM, an AIMM backfill worksheet is prepared. Outcome claims included in the worksheet are concise statements, typically accompanied by an indication of the types of outcome to be achieved. The worksheet also indicates analysis of the gap and typology, as well as the intensity and movement.

For the outcome type analysis based on outcome typologies, the *RAP* has used all the projects evaluated by IFC and validated by IEG between CY12 and the CY20 cohort for which the AIMM backfilled monitoring exercise was carried out and XPSRs were prepared. According to IFC, a sample of projects was chosen for the AIMM backfilled exercise from those projects that were active in IFC's portfolio at the time of the exercise by applying IEG's stratified sampling approach for XPSR. In this case, the projects were not selected from two separate periods, such as CY12–14 and CY17–20, to include as many projects as possible in the analysis. Instead, the projects are divided into two periods, CY12–16 and CY17–20, taking into account the distribution of similar numbers of projects in the two periods. For the CY20 cohort, only those projects whose XPSR had been validated by December 2020 were included in the analysis.

## Multilateral Investment Guarantee Agency

For MIGA, a 13-category typology of intended outcomes was developed for the *RAP 2021* report, taking advantage of MIGA's Impact Measurement and Project Assessment Comparison Tool (IMPACT) system, an ex ante assessment and monitoring tool for project outcomes that has been adapted from IFC's AIMM framework. Since IMPACT sector–specific frameworks were not available, we applied the same outcome typologies developed for IFC projects and adapted certain outcome types, particularly those included under the foreign investment dimension, to MIGA's IMPACT framework. Since MIGA has not retroactively applied IMPACT to its portfolio projects, we could not take further advantage of the IMPACT framework to determine key outcomes, including those related to foreign investment impact. Instead, we reviewed the main text of the Board proposal of MIGA projects (the President's Report) and coded descriptions of development outcomes the projects were going to achieve. We carried out analysis for all MIGA projects evaluated in FY12–14 and FY17–19 (whose validation was completed by December 2020 only). The results of the outcome claims were compared with the results presented by the PER.

Outcome type analysis faced several constraints to identifying the relationship between outcome claims and their performance for MIGA projects and the risks associated with them. First, unlike the projects of the World Bank and IFC under this *RAP* analysis, MIGA projects do not have specific ratings assigned for outcome claims. The only ratings available are those provided at dimension level for the PER, the self-evaluation report of MIGA projects. Second, before the introduction of the IMPACT framework, description of expected development outcomes was comprehensive, ranging from immediate outcomes to higher-level outcomes. Therefore, identification of specific outcome types from the broad outcome statement is challenging, and even though the coding has been done by three people covering all the outcome claims to ensure the objectivities and consistencies, there is a risk of subjective assignment of outcome types for specific outcome claims. In addition, many of those outcome claims were not accompanied by specific indicators to measure their results at approval, and most of them were not measured at evaluation.

# Methodology for Analyses on Outcome Claim Achieved and AIMM Claim Rating Change for IFC Projects

To assess the relationship between outcome types and the performance of IFC projects, two metrics were used to measure project performance: (i) change of AIMM intensity or movement rating between ex ante and monitoring, and (ii) achievement of expected outcome claims at the time of evaluation. These metrics were used because, unlike the World Bank, IFC has a self-evaluation framework that does not assign ratings at specific outcome claim level. Instead, ratings are only provided at broader development outcome dimension level and overall development outcome level.

AIMM rating change. Under the AIMM framework, the development outcome potential of projects is expressed by the combination of (i) gap and impact intensity for the project-level outcome, and (ii) market stage and market movement for market-level outcome. Since the gap and market stages of a project are set at approval and do not change during the implementation, comparison between intensity ratings at approval and at latest monitoring indicates how the actual results of the outcome claim differ from the expected results.[3] Since there are four categories under the intensity rating, the difference between the ex ante and monitoring ratings can be ranked from −4 to +4, giving opportunities to assess the variance of results per outcome type.

Outcome claims achieved. To complement the analysis of AIMM rating changes, we also assessed the extent to which expected outcome claims are achieved at evaluation by verifying the results presented at XPSR and EvNote. Since assessment of outcome claims achieved is based on the self-evaluation assessment and IEG validation, this provides an additional perspective on the analysis of the relationship between outcome types and their performance.

## Criteria to Determine Outcome Claim Achieved

The general rules are as follows:

- An outcome claim is achieved if it meets the following criteria:
  - The average score of all the indicators under one outcome claim is equal to or above 51 percent. Depending on how information is provided in the EvNote and XPSR, indicators within a project development objective are given a score ranging from 0 to 100 percent.
  - No data on the indicator are provided, but the EvNote or XPSR specifically states that the objective was met or was exceeded.
- An indicator is achieved if it meets the following criteria:
  - It meets the target set in the Board proposal and this can be measured.
  - It meets the target by 75–100 percent, with the EvNote implicitly or explicitly stating the indicator was achieved.
  - It is specifically stated in the EvNote or XPSR that the target was met, surpassed, mostly met, or will be achieved.
- An indicator is not achieved if it meets the following criteria:
  - The EvNote or XPSR specifically states the indicator was not achieved.
  - There is no target, no evidence, and no indication the indicator met the target or will meet the target in the future.
  - No outcome claim statement is provided.
  - The EvNote or XPSR states that rating or measurement of the indicator is nonconclusive.
  - No evaluation of the indicator is provided in the EvNote or XPSR.
  - The indicator was not tracked or most likely was not tracked.
  - No data or information were provided.

## Methodology for Analyses on Outcome Claim Achieved for MIGA Projects

For MIGA projects, the *RAP* followed the same approach to assess outcome claims achieved as was used for IFC projects.

## Select Analysis—World Bank

### Decomposition Analysis

The goal of the decomposition analysis is to decompose the change in average rating between two periods (in this case between FY19 and FY20) in two components—a portion that is due to changes in ratings ($r_t$) for a given category of projects *i*, and a portion that is due to changes in the share ($s_t$) of projects *i* in the overall portfolio, according to the following formula:

$$\bar{R}_t = \sum_i s_t^i \cdot r_t^i$$

$$\bar{R}_{FY20} - \bar{R}_{FY19} = \sum_i s_{FY20}^i \cdot \left(r_{FY20}^i - r_{FY19}^i\right) + \sum_i r_{FY19}^i \cdot \left(s_{FY20}^i - s_{FY19}^i\right)$$

The decomposition results are purely descriptive and should not be interpreted in a causal way. They illustrate the relative importance of composition effects (changes in share—that is, changes in the weight of certain categories of projects in the overall portfolio) versus changes in ratings for individual categories of projects. At one extreme, overall ratings can increase because a category of already highly rated projects becomes more prominent in the portfolio without the ratings of this category (or the ratings of any other category of projects) increasing from one year to the next—the increase of overall ratings would in this case be entirely attributable to changes in portfolio composition. At the other extreme, overall ratings can change because the ratings of each (or most) categories of projects increase from one year to the next, without their relative shares changing in the portfolio—the increase of overall ratings would in this case be entirely attributable to changes in ratings.

## Introduction of Novelty into Projects

This *RAP* carried out a detailed analysis of how Transport and Education GP projects have changed within a sequence of projects in a specific country. The *RAP* team chose a sample of projects that closed in FY12–14 and FY17–20 (see previous discussion of outcome type analysis). The projects were representative of the project portfolio of each GP during the two periods, at a 90 percent confidence level with a 10 percent margin of error. The team matched each sampled project to a predecessor project, defined—for the purposes of this exercise—as a project in the same GP and country that closed five years or less before the sampled successor project started, covered at least one common subsector, and had the same implementation agency as the predecessor project. In total, the sample included 75 Transport projects (of which 49 had a predecessor) and 79 Education projects (of which 55 had a predecessor). The *RAP* analyzed these projects' activities and outcome types using ICRs, Project Appraisal Documents, and ICRRs as sources of information.

The team compared predecessor and successor projects in terms of projects' IEG ratings, outcome types (as defined in the previous section), scope (measured by the number of new activities and their expansion, either into new subsectors or geographically), and the level of difficulty of new activities as determined by IEG sector experts.[4]

The team rated the novelty of each successor project as limited, moderate, or high, according to the following criteria:

» Successor projects of limited novelty had up to two new activities that the predecessor project did not have, at least one of which was assessed to be of limited difficulty, or the project scope remained roughly the same or was scaled down.

» Successor projects of moderate novelty had two to four new activities that the predecessor project did not have, at least half of which were assessed to be of moderate difficulty or had more than four new activities that were assessed to be of limited or moderate difficulty, or the project scope at least doubled in size.

» Successor projects of high novelty had more than four new activities that the predecessor project did not have, at least two of which were assessed to be of high difficulty, and the project scope at least doubled in size.

The team also examined first-time projects, which did not have predecessors. This analysis included 26 first-time projects for Transport and 24 first-time projects for Education.

## Analysis of Indicators

The analysis of indicators was based on all the objectives that, within a project sampled for the analysis of outcome types, were tagged as having the institutional strengthening outcome type.

The team analyzed each of these objectives, irrespective of whether the objective was characterized as having *only* the institutional strengthening outcome type or additional outcome types. The indicators used to measure institutional strengthening were identified, classified, and linked to the corresponding objective-level efficacy rating, based on the information provided by the ICRR.

## Target Analysis

The target analysis was based on the subsample of projects, among all those sampled for the outcome type analysis, that had objective efficacy ratings rated all high or substantial (right tail) or all negligible or modest (left tail). The reason for this selection was that because the analysis of the targets required the use of the ICR annexes as the primary source of information, it was not possible to match individual objectives (which are only defined at the ICRR stage) with efficacy ratings. The sample strategy illustrated above allowed instead for univocally assigning efficacy ratings to (the achievement of) targets. Therefore, the analysis is not generalizable to all projects.

In total, 115 projects were selected for this exercise (76 with objective efficacy ratings all rated negligible or modest, and 39 with objective efficacy ratings all rated high or substantial), for 647 project development objective indicators and targets.

The sample composition and the process to analyze targets is described in box A.2.

## Box A.2. Assessing Targets

The projects included in the analysis of targets were a subsample of all projects sampled for the *Results and Performance of the World Bank Group 2021* outcome type analysis—that is, projects with objective efficacy ratings that were all in the tails of the rating distribution. This means that projects were selected if *all* their objective efficacy ratings were either rated negligible or modest or *all* were rated substantial or high.

### Table A.2.1. Projects selected for the analysis of targets

| Project-Level Efficacy Rating at Exit | Projects (no.) | *RAP 2021* Sample (%) |
|---|---|---|
| Negligible or modest | 76 | 17 |
| High and substantial | 26 | 6 |
| High only | 13 | 3 |
| Total | 115 | 26 |

*Source:* Independent Evaluation Group.

*Note:* RAP = *Results and Performance of the World Bank Group* (report series).

The analysis was based on the information available in the indicator tables in the Implementation Completion and Results Report annexes.

For all project development objective indicators of the 115 projects (647 indicators), the following information was manually transferred into an Excel sheet: indicator name, baseline, original target, revised target, actual achievement, and comments.

The coding of the achievements was based on the following criteria:

- Not achieved: achievement of below 50 percent of the target
- Partially achieved: achievement of 50 percent to 95 percent of the target
- Fully achieved: achievement between 95 percent and 105 percent of the target
- Exceeded: achievement above 105 percent of the target
- No evidence: no or insufficient evidence on achievement
- Other1: no or unclear baseline or target

(continued)

## Box A.2. Assessing Targets (cont.)

- Other2: indicator added (shown in the column for the original indicator) or dropped (shown in the column for the revised indicator).

The available data had the following weaknesses:

- Incomplete data in the Implementation Completion and Results Report tables (for example, missing baselines or targets; missing information on achievements; empty comment sections when the reported data are not self-explanatory)
- Incorrect data (for example, an indicator speaks about a reduction in something, but the target shows an increase compared with the baseline)
- Unclear data (for example, revised targets are given in the comments section, but they are not in line with the statement in this section; from the indicator name it is not clear if a reduction or increase of the target was expected, but the target seems to indicate a reduction, the achievement shows an increase, and in the comments section it is mentioned that the indicator was fully achieved; no information is given in the actual achievement column, but in the comment section it is stated that the indicator was achieved; a revised target is shown, but it is the same as the original target; in most cases it is not clear if a target was revised or a new indicator was added)
- Unclear or missing evidence (for example, an indicator required the school management team to be judged competent in their duties, but only the number of team members who participated in training was measured)

The following rules were applied to account for some of the weaknesses:

- Missing data on targets and baselines were coded as Other1.
- However, if an indicator had no target and no evidence on the achievement, it was coded as no evidence. For example, the baseline had to be established through a study at the beginning of the project, but this study and the study to collect data on the achievement did not take place.
- All changes to indicators and targets were coded as Other2.

*Source:* Independent Evaluation Group.

## Regression Analysis

### Table A.2. Dependent Variable: Latest IEG Outcome Rating (ICRR or PPAR)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| PDO restructured (1 = yes, 0 = no) | −0.344*** | −0.347*** | −0.362*** | −0.525*** |
| Log (project volume) | 0.171*** | 0.165*** | 0.168*** | 0.227*** |
| Human Capital Index 2020 | 1.495*** | 1.460*** | 1.543*** | 1.722* |
| Log (days between project closing and ICRR completion) | −0.576*** | −0.589*** | −0.539*** | −0.748*** |
| FY20 (1 = yes, 0 = no) | | 0.323*** | 2.433*** | 0.253* |
| PDO restructured x FY20 | | | 0.125 | |
| Log (project volume) x FY20 | | | −0.012 | |
| Human Capital Index 2020 x FY20 | | | −0.343 | |
| Log (days between project closing and ICRR completion) x FY20 | | | −0.322 | |
| Constant | 6.1470*** | 6.205*** | 5.860*** | 6.805*** |
| TTL as fixed effect | No | No | No | Yes |
| Observations | 938 | 938 | 938 | 938 |
| *R*-squared | 0.114 | 0.129 | 0.130 | 0.126 |

*Source:* Independent Evaluation Group.

*Note:* ICRR = Implementation Completion and Results Report Review; IEG = Independent Evaluation Group; PDO = project development objective; PPAR = Project Performance Assessment Report; TTL = task team leader; FY = fiscal year.
***p < .001.

**Table A.3.** Dependent Variable: Latest IEG Efficacy Rating (Individual Objective-Level) Logit Regression Coefficients

| | (1) | (2) |
|---|---|---|
| Access to services | 0.501** | 0.301 |
| Quality of services | 0.150 | 0.114 |
| Public assets | 0.491 | 0.885 |
| Natural capital | −0.259 | −0.215 |
| Use of services | 0.019 | 0.016 |
| Temporary income | 0.767 | −0.042 |
| Awareness and attitudes | 0.257 | 0.134 |
| Human capital | 0.761* | 0.312 |
| Employability | −0.04 | 0.207 |
| Citizen engagement | 0.361 | 0.746 |
| Legal or regulatory | 0.166 | 0.139 |
| Institutional capacity | −0.075 | −0.078 |
| Accountability | 0.249 | 0.089 |
| Enterprise or sector performance | 0.881*** | 0.736** |
| Sector expansion | 0.320 | 0.925* |
| Equity or inclusion | 0.683** | 0.453 |
| M&E quality | | 1.580*** |
| Constant | 0.402 | −3.001*** |
| Observations | 968 | 968 |
| Pseudo *R*-squared | 0.0287 | 0.1643 |

*Source:* Independent Evaluation Group.

*Note:* IEG = Independent Evaluation Group; M&E = monitoring and evaluation.
*p < .05 **p < .01 ***p < .001.

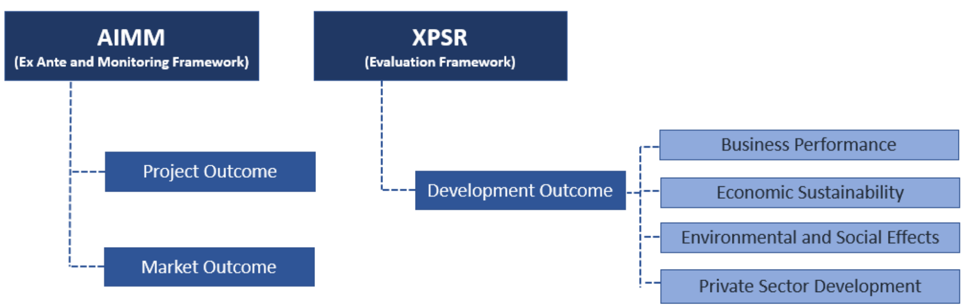
# Select Analysis—International Finance Corporation

## Outcome Potential Analysis

An analysis of the relationship between outcome potential and XPSR ratings was carried out, reviewing the scores and ratings for development outcome from AIMM and XPSR to test this *RAP*'s original hypothesis that projects intended to address higher levels of development challenges have lower performance ratings or larger variance in performance. AIMM ratings consider ex ante the level of development challenges to be addressed by the projects (ex ante ratings), whereas the XPSR ratings provide evaluative assessment after implementation (the ex post or performance ratings). By comparing potential AIMM ratings for the sampled projects and their corresponding XPSR ratings for the equivalent dimensions (economic sustainability and private sector development), we can assess the relationship between a project's development potential versus its risks or performance expressed and their variance. We can test whether projects with higher potential (higher AIMM rating or score) at these dimensions have more variance in XPSR ratings.

The AIMM framework is composed of two dimensions (project outcome and market outcome) and the XPSR framework is composed of four dimensions (project business performance, economic sustainability, environmental and social effect, and private sector development; figure A.7). There are two XPSR dimensions, economic sustainability and private sector development, that are corresponding to the project outcome and market outcome dimensions in AIMM.

Figure A.7. Comparison of the Development Outcome Dimensions under AIMM Framework and XPSR Framework




*Source:* Independent Evaluation Group.

*Note:* AIMM = Anticipated Impact Measurement and Monitoring; XPSR = Expanded Project Supervision Report.

For instance, outcomes on specific stakeholders—such as customers, employees, governments, suppliers or distributors, and overall communities—are considered in the AIMM's project outcome dimensions and economic sustainability under XPSR. Likewise, beyond-the-project outcomes at the market level that are related to sector and industry are included in the

AIMM's market outcome and private sector development dimensions in the XPSR. The AIMM and XPSR may use at times different metrics to measure similar effects. For instance, AIMM can use gross value added to assess economywide effect as a part of project outcome and XPSR looks at economic rate of return to assess overall benefits for the society in a quantitative manner, in addition to effects on immediate beneficiaries under its economic sustainability dimension.

However, there are differences in how the outcomes are assessed by these two systems, AIMM and XPSR. First, the AIMM potential rating is provided based on the magnitude of a country's development challenge (expressed by gap or stage) and degree of the project's contribution (expressed by intensity or movement) to addressing such a gap. This AIMM potential rating is then discounted by the likelihood of the project achieving such an outcome. Ex ante AIMM ratings reflect the expected level of outcome, but the monitoring AIMM rating is based on the actual results. XPSR, however, assesses the achievement of expected development outcome targets set by IFC at project approval. In this regard, an AIMM rating consists of a judgment about the *absolute* impact potential of a project, whereas an XPSR rating expresses a judgment about the *relative performance* of a project against IFC's original expectations.[5] In addition, XPSR also assesses performance of the projects in comparison with the industry benchmarks and peers, considering broader market context. Finally, AIMM focuses on a limited number of key (intended) development outcomes under both project outcome and market outcome dimensions, whereas XPSR takes into account a large number of indicators considering various elements of development outcome, including unintended outcomes (positive or negative). XPSR also includes an assessment on projects' business performance (financial sustainability) and environmental and social aspects (environmental sustainability) for a project's overall development outcome.

## Project-Level Analysis

At the project level, it was also tested whether projects intended to have higher market impact suffered lower ratings or higher variance (meaning higher risk). To distinguish projects with the intention to achieve higher market impact, sampled projects were divided into three groups: (i) projects with expected market outcome claims with indicators, (ii) projects with expected market outcome claims without indicators, and (iii) projects without market outcome claims. To assess the overall level of performance at evaluation for each group, the average XPSR development outcome scores were calculated by assigning points for each development outcome rating category (highly successful: 6, successful: 5, mostly successful: 4, mostly unsuccessful: 3, unsuccessful: 2, highly unsuccessful: 1). Then the actual development outcome ratings for each category were compared with the variance of the XPSR ratings. In addition, ex ante AIMM scores at the project level were identified for each group as an indication of their outcome potential. For more on the concept of project potential, please see the section Outcome Potential Analysis on page 107.

[1] The analysis was based on the content of the Efficacy section of Implementation Completion and Results Report Reviews. In each review, the validator structures the discussion of achievement of objectives (efficacy) according to the project development objective, creating a separate section for each objective inherent within the project development objective. Each objective-level section contains (i) a title that states the objective, (ii) a narrative explanation of the extent of the achievement (including a summary of the intended theory of change, outputs, outcomes, and the summary of key evidence that supports the achievement claims), and (iii) a rating of achievement. For the analysis of outcome types, the team reviewed each objective-level title and narrative then assigned one or more outcome types that fit the information presented about what the project intended to accomplish with respect to that specific objective.

[2] The fiscal years (FY)17–20 Q2 cohort sample was composed of projects that closed in FY17, FY18, FY19, and the first two quarters of FY20 that had Implementation Completion and Results Report Reviews completed, with ratings available in World Bank systems as of January 21, 2021.

[3] Since impact intensity and market movement are provided based on the rating range, if the actual results fall in the same range, the ratings do not change.

[4] See endnote 16 of chapter 2 for more details.

[5] When judging achievements against targets, comparison between ex ante Anticipated Impact Measurement and Monitoring ratings and monitoring Anticipated Impact Measurement and Monitoring ratings would reflect outcome achievements within a target range, and Expanded Project Supervision Report ratings tend to reflect outcome achievements against point estimates.

# Appendix B. World Bank Outcome Type Analysis and Classifications

## Background

The goal of the outcome type analysis is to describe the objectives of World Bank projects (in terms of intended development outcomes) and assess the relationship between those outcomes and the objective-level efficacy ratings. The hypothesis to be tested was that some types of outcomes may be more challenging than others and therefore may be associated with higher or lower outcome efficacy ratings (they may have different average probabilities of success, or a wider variance of ratings). This analysis expands on the *Results and Performance of the World Bank Group 2020* (*RAP 2020*) report, which distinguished among 4 outcome *levels*, ranging from outputs to early outcomes, intermediate outcomes, and long-term outcomes, and proposes 16 outcome *types*. The methodology to construct the outcome types is described in appendix A.

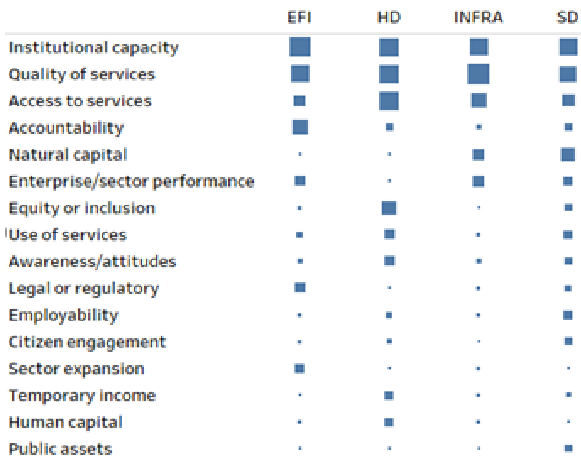
## Outcome Types Description

The three most common outcome types at project level, in order of frequency, are (i) enhanced institutional capacity, (ii) improved quality of services, and (iii) expanded access to services. Each outcome type comprises a wide variety of activities.

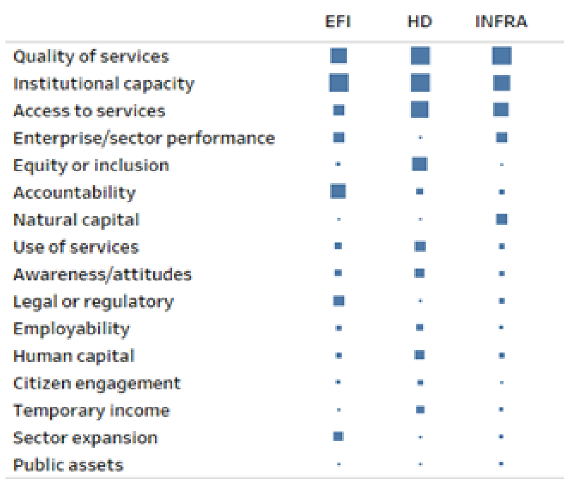
In every Practice Group, enhanced institutional capacity and improved service quality are the most common outcome types, included in 62.8 and 62.4 percent of World Bank projects and in 39 and 38 percent of project objectives. Expanded access to services is the next most frequent outcome type at project level in Human Development, Infrastructure, and Sustainable Development, but not in Equitable Growth, Finance, and Institutions, where enhanced accountability, transparency, and governance play a more important role. Equity and inclusion—a cornerstone of achieving the Bank Group's shared prosperity goal—is the sixth most frequent outcome type at project level, mostly prevalent in Human Development and, to some extent, in Sustainable Development (see figure B.1).

**Figure B.1.** Distribution of Outcome Types among the World Bank's Four Practice Groups

a. Outcome types as percentage of all objectives




b. Projects tagged with specific outcome types (%)




*Source:* Independent Evaluation Group.

Note: The size of the squares corresponds to the percentage of objectives characterized by each outcome type. One objective may be characterized by several outcome types, so the percentages would add up to more than 100 percent. EFI = Equitable Growth, Finance, and Institutions; HD = Human Development; INFRA = Infrastructure; SD = Sustainable Development.

Each outcome type comprises a wide variety of activities of different durations, intensities, and challenges. Even projects with very similar objectives and outcome types can have very different activities. A project can achieve expanded access to services through, for example, the construction of all-weather roads or the provision of microcredits. Table B.1 shows the different types of activities that compose four of the most important outcome types. These activities can be of different durations and intensities as well. For instance, within enhanced institutional capacity, the most frequently found activities are aimed at skills development, such as training, workshops, or study tours, and these often have short-term impacts. However, activities aimed at streamlining organizational structures also have a long-term objective, such as creating new government agencies, establishing sustainable funding streams, and enhancing government systems. Activities that expand access to services often build infrastructure. They may provide scholarships or credits, which are very different in nature. Activities that improve the quality of services more often rehabilitate and maintain infrastructure, but they may also include training or system development, showing an overlap with enhanced institutional capacity. Meanwhile, activities that enhance equity and inclusion for vulnerable populations frequently provide financial support or other goods to increase equitable access to services and infrastructure. Again, they may also provide systems or training.

**Table B.1.** Activities within Outcome Types

| Outcome Types | Key Activity Types |
|---|---|
| Institutional strengthening | Skill development |
| | Development of new or enhanced systems |
| | Streamlining organizational structures |
| | Reviews, assessments, and audits |
| | Development of action plans, policies, and strategies |
| | Development of guidelines, handbooks, and manuals |
| | Preparation of standards, agreements, or legal frameworks |
| | Human resource development |
| | Operational assistance |
| Access to services | Construction of infrastructure |
| | Supply of equipment or goods |
| | Provision of scholarships, financial aid, school fee subsidies, and feeding programs |

*(continued)*

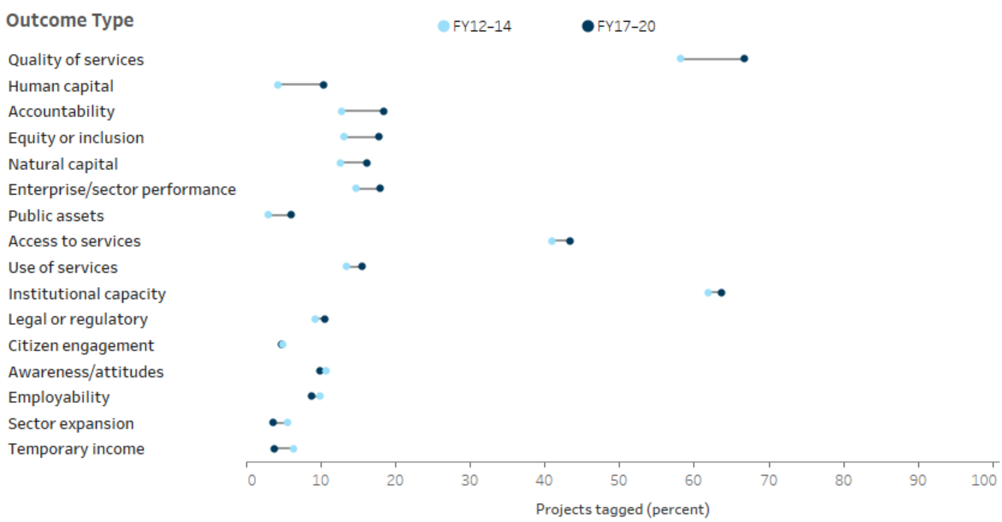
| Outcome Types | Key Activity Types |
|---|---|
| | Provision of credit lines, microcredits, and longer-term funding |
| Service quality | Rehabilitation and maintenance of infrastructure |
| | System installation and strengthening and process and procedure simplification |
| | Staff training and professional development |
| | Development of curricula and learning materials, pedagogical methods, policies and standards, performance appraisal methods, operations manuals, and quality assurance frameworks |
| Equity and inclusion | Studies or advice to improve targeting |
| | Improvement of targeting processes |
| | Creation of monitoring or other systems for targeting |
| | Scholarships, grants, loans, subsidies, and cash transfers |
| | Implementation of program targeted services to specific populations (for example, girls) |
| | Development of guides and modules, policies, and work plans |
| | Training and awareness campaigns |
| | Creation of governance bodies |

Source: Independent Evaluation Group.

## Changes in Outcome Types

Since 2012, the number of outcome types in World Bank projects has increased, and their relative importance has changed. The average number of outcome types has increased from 2.8 per project in fiscal year (FY)12–14 to 3.2 per project in FY17–20. Moreover, some outcome types have become more prevalent than others. Improved service quality, increased human capital, enhanced equity and inclusion, and enhanced accountability, transparency, or governance increased the most from the earlier to the later period (figure B.2). Meanwhile, temporary relief to individuals, expanded productive sector, improved individual employability or livelihood, and changed awareness, attitudes, or behaviors all became less common from the earlier to later periods.

Figure B.2. Change in Frequency of Outcome Types, FY12–14 to FY17–20




*Source:* Independent Evaluation Group.

*Note:* Because each project can have more than one outcome type, the prevalence of the outcome types in FY12–14 and in FY17–20 does not total 100. FY = fiscal year.

The relative importance of outcome types changed within Practice Groups between FY12–14 and FY17–20Q2. The significant overall increase in improved service quality largely came from the 25 percentage point increase in Sustainable Development and the 11 percentage point increase in Infrastructure projects from FY12–14 to FY17–20Q2. Expanded access to services also increased in these two Practice Groups, but to a lesser extent (+7 and +8 percentage points). Enhanced institutional capacity had always been prevalent across all Practice Groups, but it increased substantially in Sustainable Development and Human Development projects. Human Development projects also saw large increases in the increased human capital and enhanced equity and inclusion outcome types, with human capital increasing by 18 percentage points and equity and inclusion by 15 percentage points. Equitable Growth, Finance, and Institutions projects, for their part, increased the enhanced accountability, transparency, or governance outcome type by 11 percentage points between the two periods.

The prevalence of certain outcome types changed because of a change in the types of projects undertaken, based on an analysis of the projects of two Global Practices (GPs). An analysis of the Education GP's portfolio indicates that projects that closed in the earlier period (FY12–14) focused relatively more on early childhood education and basic and secondary education. Meanwhile, projects that closed in the later period (FY17–20) focused relatively more on higher education—including tertiary education and vocational and technical education. This switch meant the later period had more outcomes than the earlier period related to improved individual employability and livelihood, increased human capital, and enhanced accountability, transparency, and governance. In the Transport GP during FY17–20, more projects focused on access to services and institutional strengthening than in the previous period—for example, by enhancing the institutional capacity to better manage roads and other transport activities. In addition, as road safety and decarbonization became greater priorities over the past decade, increased human capital and natural capital outcome types became more common in Transport's portfolio.

The prevalence of outcome types in different GPs changed not only because the project focus changed but also because what projects measured changed. Over the past decade, several outcomes have become greater priorities to measure, which was reflected in changing outcome types. For example, during FY17–19, Education projects collected more evidence than in the earlier period on skills development and employment (quite likely in response to an increased focus on employability), which explains the increase in the corresponding outcome types. For instance, early childhood education projects that closed in the later period included more indicators on learning, which raised the frequency of increased human capital outcomes in early childhood education projects. In another example, road safety indicators became a greater focus in Transport projects, leading to more frequent human capital outcomes in that GP as well.[1]

## Outcome Types and Efficacy Ratings

Four outcome types are more likely than the others to have higher objective-level efficacy ratings, but this is mostly because of the higher monitoring and evaluation (M&E) quality of the projects that pursue these outcome types. This *RAP* carried out a regression analysis that controlled for outcome types and other project characteristics. It found that four outcome types—expanded access to services, increased human capital, improved enterprise and sector performance, and enhanced equity and inclusion—are more likely to achieve higher objective-level efficacy ratings than other outcome types. However, the relationship among the four outcome types and the efficacy rating is largely due to the M&E quality of the project. Higher M&E quality also has a very strong independent effect on efficacy ratings—in line with what is observed in the analysis of project outcome ratings—indicating that, no matter what the outcome type, good M&E matters in achieving higher ratings.[2]

## Outcome Types for World Bank Projects: Coding Matrix

Table B.2 lists all the outcome types identified for this year's *RAP* (refer to table B.1 for a description of each category).

Table B.2. Codes for Objective-Level Assessment of Outcome Types

| Outcome Type (Name) | Definition |
|---|---|
| 01 full<br>Access to services expanded<br>01 short<br>Access to services | The intended change described in the PDO or key indicators (or outputs or activities) is about expanded access to basic public services.<br>Here we define "access" as the opportunity of more project beneficiaries (people) to benefit from or use a basic public service (or service system).<br>An increase in access could happen through providing more of a service, or bringing a service closer to people, making a service more affordable, or making other changes that enable more people to access a service.<br>Examples of basic public services include transport (through roads, bridges, ports, and so on), energy, water, education, health care, and so on. |

*(continued)*

| Outcome Type (Name) | Definition |
|---|---|
| 02 full<br>Quality of services improved<br>02 short<br>Quality of services | The intended change described in the PDO or key indicators (or outputs or activities) is about improving or upgrading the quality of services provided. Quality may be understood as reliability, timeliness, cleanliness, frequency, and so on. The intended change may have to do with improvement or maintenance of the quality of services. Project beneficiaries have the opportunity to benefit from or use a better system or service.<br><br>Sometimes there is a fine line between access to services and quality of services. For instance, if the project replaces electricity from solar energy with electricity from the grid, it might be considered access or quality of service depending on the counterfactual (no electricity—which would point to access to services; or no electricity from the grid—which would point to quality of services). In such cases, we pay special attention to the objective statement and indicators to take a decision. |
| 03 full<br>Public assets improved<br>03 short<br>Public assets | The intended change described in the PDO or key indicators (or outputs or activities) is about improving or enhancing the condition of public assets rather than public basic services. By public assets, we mean built assets rather than natural assets.<br><br>We understand improving public assets as building, repairing, or improving structures or spaces that contribute to a better quality of life but are not directly connected to a system or network that provides basic services. Examples of public assets include museums, urban spaces, and monuments.<br><br>Improving cultural assets and heritage sites goes under this code.<br><br>If the intended change involves building or improving roads, bridges, ports, schools, hospitals, and so on, then use either access to services or quality of services. Our understanding in such cases is that construction or improvement of structures or assets that enable or support the provision of basic public services is an intermediate step toward expanding access to services or improving the quality of services. |
| 04 full<br>Natural capital sustained<br>04 short<br>Natural capital | The intended change described in the PDO or key indicators (or outputs or activities) is about improving, enhancing, or preserving the condition of natural assets. *Natural capital* here is understood as a general term for natural assets. We use the term *natural capital* for this outcome type for contrast with *human capital* (below).<br><br>Natural assets may include air, water, forests, minerals, land, biodiversity, or other aspects of the natural environment.<br><br>Improving, enhancing, or preserving the condition of natural assets may be direct (such as cleaning up a waterway or replanting a deforested area) or indirect (such as changes to energy-saving measures intended to preserve or improve the natural environment). |

*(continued)*

| Outcome Type (Name) | Definition |
|---|---|
| 05 full<br>Use of services or assets increased<br>05 short<br>Use of services | The intended change described in the PDO or key indicators (or outputs or activities) is about creating or enhancing incentives for people to use basic services, goods, technologies, or other kinds of assets. Increased use may involve adequate prices, regulation, incentives, or information about the existence or benefits of a service, good, technology, or other kind of asset.<br><br>We distinguish use of services from access to services or quality of services by considering what is happening with incentives. With use of services, the goal is not (only) to provide more of a service or a better service but also to incentivize its use. We look for actions on both the demand side and the supply side. |
| 06 full<br>Temporary relief to individuals provided<br>06 short<br>Temporary income | The intended change described in the PDO or key indicators (or outputs or activities) is about increased benefit to individuals through temporary income support, such as cash transfers, in-kind support, or public work schemes. Temporary income support may take the form of social safety nets. Temporary income support is sometimes called "temporary relief."<br><br>The intent of this code is to capture cash support or in-kind support that is dependent on financing and, hence, temporary in nature. This code can be contrasted with code 9, employability or livelihoods, which is intended to capture progress in terms of sustainability of income or livelihood.<br><br>The following initiatives also go under this code.<br>» All conditional cash transfers, that is, cash transfers with conditions tied to participation in education, regularly visiting health care facilities, or receiving vaccinations.<br>» Provision of short-term employment during project implementation. Public work schemes go here, as well as, for example, labor-intensive road building.<br>» In-kind support such as food stamps.<br>When a project provides skills, loan financing, or other support meant to increase the employability of the beneficiaries in the medium and longer term, code this under "Individual employability or livelihood improved." |

*(continued)*

| Outcome Type (Name) | Definition |
|---|---|
| 07 full<br>Awareness, attitudes, or behaviors changed<br>07 short<br>Awareness and attitudes | The intended change described in the PDO or key indicators (or outputs or activities) is about increased awareness, improved attitudes, or modified behaviors on the part of project beneficiaries to achieve better and more sustainable results.<br><br>The project beneficiaries are individuals, such as citizens, customers, users, or employees. Awareness, attitudes, and behaviors have to do with bringing benefit to individuals or to a population, such as awareness of handwashing for rural populations, or awareness of the benefits of not smoking, or behaviors related to recycling in a public institution.<br><br>If the awareness creation is for members of an institution to improve the way they carry out their function, the change refers to institutional strengthening.<br><br>Note that this code is likely a "qualifier," that is, not necessarily the explicit final step in the project's theory of change. This code may therefore be applied as a secondary code. This outcome type code was included because of the increasing emphasis of the World Bank Group on influencing or incentivizing modified attitudes or behaviors for greater sustainability of interventions. |
| 08 full<br>Human capital increased<br>08 short<br>Human capital | The intended change described in the PDO or key indicators (or outputs or activities) is about increasing human capital, which includes increasing the knowledge or skills of individuals for their own benefit, or improving their health status, or other increases in human capital.[a]<br><br>However, code other types of substantial training or education as follows.<br><br>» Training of teachers, health workers, or other basic service providers—quality of services.<br>» Training of employees of a ministry or other central government institution—institutional capacity.<br><br>For training of staff to carry out project activities, do not use this human capital code. Instead, use the code appropriate for the project activity or objective that the staff are expected to accomplish using the skills or knowledge they acquire through training. |

*(continued)*

| Outcome Type (Name) | Definition |
|---|---|
| 09 full<br>Individual employability or livelihood improved<br>09 short<br>Employability | The type of change described in the PDO or key indicators (or outputs or activities) is about increasing the employability of individuals and their ability to increase their earnings in a sustainable way, that is, beyond the term of the project.<br><br>This includes fostering micro enterprises or promoting the development of small (family) business to provide livelihoods for individuals and thus enhance earnings. Examples include support to self-employed entrepreneurs to prepare business plans and provision of small grants or loans. Note that small loans in this category should be for business purposes, not for personal consumption or living expenses.<br><br>The emphasis is on providing individuals or individually owned very small, often informal, businesses with the means to strengthen their position in employment (self-employment or paid employment) to ensure better access to employment—or access to better employment—and higher earnings. As such, the focus in this code is on improving the livelihoods of individuals.<br><br>Do not use this code for public work schemes or short-term employment generation only related to project implementation. Instead, such initiatives should be coded to code 6, temporary income. |
| 10 full<br>Citizen engagement, or community participation enhanced<br>10 short<br>Citizen engagement | The type of change described in the PDO or key indicators (or outputs or activities) is about setting up mechanisms that give beneficiaries, users, or citizens a "say" or a "voice," for either intrinsic or instrumental value.<br><br>Use this code only if the project has a substantial component for supporting, promoting, or incentivizing citizen engagement—it should go beyond stakeholder consultations or workshops.<br><br>Note that similar to code 16, equity or inclusion, and code 7, awareness, attitudes, or behaviors, this code is generally a "qualifier" and hence often used as an additional code. |

*(continued)*

| Outcome Type (Name) | Definition |
|---|---|
| 11 full<br>Legal or regulatory context improved<br>11 short<br>Legal or regulatory | The type of change described in the PDO or key indicators (or outputs or activities) is about creating or improving regulations, laws, codes, working conditions, standards, and environmental requirements.<br><br>For this code, we look for policies, regulations, and so on that plausibly would lead to broader changes at the country level or across a population or sector. To be coded here, the creation of a legal framework should be the final outcome or a significant change in the theory of change leading to achievement of the intended objective.<br><br>Changes in laws, regulations, or procedures related to trade facilitation, doing business, or improvements to investment climate should be coded here if the change is the final outcome in the theory of change.<br><br>If the purpose of the new or improved laws or other legal instruments is to create authorities, institutions, or funding mechanisms (for example, a road fund), do not use this code; code these instead under code 12, institutional capacity.<br><br>Reflect on the difference between "Legal or regulatory context improved" and "Capacity of institutions to perform institutional functions enhanced." |

*(continued)*

| Outcome Type (Name) | Definition |
|---|---|
| 12 full<br>Capacity of institutions to perform institutional functions enhanced<br>12 short<br>Institutional capacity | The type of change described in the PDO or key indicators (or outputs or activities) is about increasing the capacity of public and private institutions to better carry out their institutional function in a sustainable way (that is, in a plausibly permanent way, for the longer term beyond the duration of the project).<br><br>If the project strengthens institutional capacity with a purpose beyond that capacity, this code usually becomes the secondary code, and the purpose of institutional strengthening determines the primary code, as follows.<br><br>» When the aim of institutional strengthening is to enhance the performance of firms, sectors, or markets—either explicitly (such as when performance of firms is denoted in the objective statement) or implicitly (such as when enhanced performance of firms, sectors, or markets is measured through greater productivity, more efficiency, or more revenues)—use code 14, enterprise or sectoral, as a primary code and use code 12, institutional capacity, as a secondary code.<br><br>» When the aim of institutional strengthening is to enhance the accountability, transparency, or governance of public administration—either explicitly (such as when accountability of public administration is denoted in the objective statement) or implicitly (such as when enhanced performance of public administration is measured, or when or citizens are intended to benefit from more transparency or less corruption)—use code 13, accountability, as a primary code, and use code 12, institutional capacity, as a secondary code.<br><br>» When the aim of institutional strengthening is to enhance performance in delivering services—either explicitly (such as when improved delivery of services is denoted in the objective statement) or implicitly (such as when enhanced performance in delivering services is measured, for example, by fewer days needed to obtain a construction permit)—use code 02, quality of services, as a primary code and use code 12, institutional capacity, as a secondary code.<br><br>» When the aim is institutional strengthening itself and only outputs are measured (for example, staff trained, reports produced, system implemented), use code 12, institutional capacity, as a primary code. |

*(continued)*

| Outcome Type (Name) | Definition |
|---|---|
| 13 full<br>Accountability, transparency, or governance enhanced<br>13 short<br>Accountability | The type of change described in the PDO or key indicators (or outputs or activities) is about increasing transparency, accountability, and openness and combating corruption. These initiatives can contribute to increased trust, greater participation, and more inclusion. In many cases, such initiatives may also aim at promoting efficient government and effective service delivery, facilitating private sector growth, building resilient institutions, or earning the confidence of citizens. |
| 14 full<br>Enterprise or sectoral performance improved<br>14 short<br>Ent or sector performance | The type of change described in the PDO or key indicators (or outputs or activities) is about improved productivity, efficiency, profitability, or competitiveness of a firm (including state-owned enterprises) or of a sector within the country economy.<br>Improvements in enterprise or sectoral performance might involve changes in production aspects (for example, the ways the entity or sector operates), organizational structure (for example, staff reorganization), use of technologies, reorganization of value chains, fostering innovation, or other improvements to productivity.<br>This code focuses on the performance of firms or sectors of the economy. Improvements in the functioning of markets may also be coded here.<br>When the aim is enhanced individual employability or household livelihoods, use code 09, employability. |
| 15 full<br>Productive sector expanded<br>15 short<br>Sector expansion | The type of change described in the PDO or key indicators (or outputs or activities) is about increasing the dimension or size of a productive sector within the country economy. The purpose of change may be to expand a sector by, for example, supporting start-ups or higher investment in firms. In such cases the aim would be increases in both capital and employment. Activities might be structured to create the conditions for productivity, with the aim of increasing the size of the relevant productive sector.<br>This code focuses on initiatives aimed at the growth or expansion of a sector.<br>When the aim is improving the productivity or competitiveness of firms or of a sector, use code 14, ent. or sector performance.<br>When the aim is enhanced individual employability or household livelihoods, use code 09, employability. |

*(continued)*

| Outcome Type (Name) | Definition |
|---|---|
| 16 full<br>Equity or inclusion enhanced<br>16 short<br>Equity or inclusion | The type of change described in the PDO or key indicators (or outputs or activities) is about achieving greater inclusion of, or more equitable outcomes for, groups that experience disadvantage based on gender, ethnicity, disability, or socioeconomic status, including poverty status.<br>This code aims to capture proactive measures adopted by the project to achieve greater equity (such as quotas, incentives, components designed explicitly for specific "disadvantaged" groups, and so on).<br>Although all Bank Group projects are, in principle, about decreasing poverty, the focus of this code is on disruption to dynamics of exclusion that may exist. A few rules are applied in deciding to use this code, as follows.<br>» Focus on explicit and proactive measures aimed at promoting inclusion or discouraging exclusion.<br>» Targeting based on poverty status (for example, a project implemented in a poor area; a project that applies a qualifying condition based on income) is excluded, that is, the presence of targeting within a project is not sufficient to apply this code.<br>» Increasing the capacity of institutions to identify and target disadvantaged groups is included.<br>» Tracking participation of disadvantaged groups (for example, reporting the percentage of women benefiting from an intervention) is excluded, that is, tracking participation is not sufficient to apply this code.<br>This code is likely to be a "qualifier" of another outcome type; hence it is often used as an additional code. |

*Source:* Independent Evaluation Group.

*Note:* PDO = project development objective.
a. For more information, visit the Frequently Asked Questions section of the World Bank's Human Capital Project website at https://www.worldbank.org/en/publication/human-capital/brief/the-human-capital-project-frequently-asked-questions#1.

[1] The *Results and Performance of the World Bank Group* team identified outcome types by looking at projects' objective statements, indicators, and activities.

[2] For some outcome types (improved enterprise or sectoral performance; enhanced equity and inclusion; changed awareness, attitudes, or behaviors; and enhanced accountability, transparency, and governance), higher or lower monitoring and evaluation quality makes a small difference to efficacy ratings with respect to other outcome types. For two other outcome types (increased human capital and improved legal and regulatory context), the effect of monitoring and evaluation quality is instead large.

# Appendix C. International Finance Corporation and Multilateral Investment Guarantee Agency Outcome Type Analysis and Classifications

## International Finance Corporation Outcome Type Analysis

The International Finance Corporation (IFC)'s project outcomes fall into two broad categories—project-level claims and market-level claims (box C.1). Project-level claims, or outcomes, are defined as a project's direct and indirect effects on stakeholders, the economy, and the environment. Market claims are derived effects, defined as a project's ability to catalyze systemic changes beyond those effects brought about by the project itself. The IFC 3.0 strategy explicitly prioritizes "creating markets," which falls into the market-level category. Overall, all IFC projects had project-level claims and 86 percent had market-level claims.

### Box C.1. Examples of Project and Market-Level Indicators

The following are examples of the outcome types.

Financial Institutions Group projects: amount of outstanding small and medium enterprise loans, share of microfinance loans, number of automated teller machines, reduced nonperforming loans in the total portfolio, demonstration of the viability of lending to small and medium enterprises in a country.

Infrastructure projects: power generated, number of airport service users, number of passengers with access to the road, diversification of energy mix.

Manufacturing, Agribusiness, and Services projects: number of cattle suppliers, purchase from domestic suppliers, number of students enrolled, demonstration effect on the local agribusiness industry, establishing viability of green buildings, and promoting replication.

Disruptive Technologies and Funds projects: percentage of companies with growth in revenue and earnings before interest, taxes, depreciation, and amortization; facilitation of investee companies' emergence as regional players.

*Source:* Independent Evaluation Group.

Most IFC projects pursue two outcome types: improved access to goods and services and increased market competition. Overall, the *Results and Performance of the World Bank Group* report identified 13 outcome types with several subcategories for IFC projects (see table C.1). These outcome types are aligned with those defined by the Anticipated Impact Measurement and Monitoring (AIMM) sector framework. For the entire period of calendar years 2012–20, 68 percent of IFC projects focused on improving access to goods and services and 53 percent on increasing market competitiveness. Meanwhile, other important outcome types are less common, including quality or affordability of goods and services, improved living standards of individuals, and improved sales and profitability of enterprises.

**Table C.1.** Share of IFC Investment Projects with Certain Outcome Types (percent)

| Outcome Type | 2012–16 (*N* = 119) | 2017–20 (*N* = 117) | 2012–20 (*N* = 236) |
|---|---|---|---|
| Project-level outcome | | | |
| 1.1 Access to goods and services (1.1.1–1.1.3) | 66 | 70 | 68 |
| 1.1.1 Access to goods and services (MSME) | 30 | 33 | 32 |
| 1.1.2 Access to goods and services (female) | 2 | 10 | 6 |
| 1.1.3 Access to goods and services (customers) | 39 | 42 | 40 |
| 1.2 Quality/affordability of goods and services | 8 | 6 | 7 |
| 1.3 Enhanced capacity of final beneficiaries | 18 | 9 | 13 |
| 1.5 Improved sales/profitability of enterprises | 22 | 12 | 17 |
| 2.1 Suppliers/distributors reached | 6 | 15 | 10 |
| 2.3 Improved sales/profitability of suppliers/distributors | 9 | 15 | 12 |
| 3.1 Increased employment | 19 | 28 | 24 |
| 4.1 Increased transfers to the government | 4 | 3 | 4 |
| 6.2 GHG reduction | 13 | 13 | 13 |
| Market-level outcome | | | |
| 9. Competitiveness in the market | 52 | 55 | 53 |
| 10. Resilience in the market | 9 | 11 | 10 |
| 11. Integration in the market | 14 | 9 | 12 |
| 12. Inclusiveness in the market | 8 | 5 | 6 |
| 13. Sustainability in the market | 6 | 5 | 6 |

*Source:* Independent Evaluation Group.

*Note:* Individual projects can have multiple outcome types. GHG = greenhouse gas; IFC = International Finance Corporation; MSME = micro, small, and medium enterprise.

IFC's specific industry groups tend to support certain outcome types (table C.2). Its four industry groups are Disruptive Technologies and Funds; Financial Institutions Group; Infrastructure; and Manufacturing, Agribusiness, and Services. At project outcome level, the Financial Institutions Group and

Infrastructure industry groups frequently implement projects for access, whereas Manufacturing, Agribusiness, and Services frequently implements projects for suppliers and distributors. All industry departments support market outcomes, but many Financial Institutions Group projects also enhance sustainability in the market, and many Infrastructure projects enhance resilience in the market.

**Table C.2.** Share of IFC Investment Projects with Certain Outcome Types per Industry Groups (percent)

| Outcome Type | 2012–16 (*N* = 119) | | | | 2017–20 (*N* = 117) | | | |
|---|---|---|---|---|---|---|---|---|
| | CDF | FIG | INR | MAS | CDF | FIG | INR | MAS |
| Project-level outcome | | | | | | | | |
| 1.1 Access to goods and services (1.1.1–1.1.3) | 8 | 30 | 18 | 9 | 3 | 37 | 17 | 13 |
| 1.1.1 Access to goods and services (MSME) | 8 | 22 | 0 | 0 | 3 | 28 | 0 | 3 |
| 1.1.2 Access to goods and services (female) | 0 | 2 | 0 | 0 | 0 | 10 | 0 | 0 |
| 1.1.3 Access to goods and services (customers) | 0 | 11 | 18 | 9 | 0 | 12 | 17 | 12 |
| 1.2 Quality/affordability of goods and services | 0 | 0 | 4 | 3 | 0 | 0 | 2 | 3 |
| 1.3 Enhanced capacity of final beneficiaries | 17 | 0 | 0 | 1 | 6 | 1 | 0 | 2 |
| 1.5 Improved sales/profitability of enterprises | 19 | 1 | 1 | 2 | 10 | 2 | 0 | 1 |
| 2.1 Suppliers/distributors reached | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 14 |
| 2.3 Improved sales/profitability of suppliers/distributors | 0 | 0 | 1 | 8 | 0 | 1 | 0 | 14 |
| 3.1 Increased employment | 0 | 0 | 8 | 11 | 0 | 0 | 7 | 22 |
| 4.1 Increased transfers to the government | 0 | 0 | 2 | 3 | 0 | 0 | 3 | 1 |
| 6.2 GHG reduction | 0 | 0 | 10 | 3 | 1 | 0 | 9 | 3 |
| Market-level outcome | | | | | | | | |
| 9. Competitiveness in the market | 9 | 20 | 13 | 11 | 9 | 23 | 12 | 11 |
| 10. Resilience in the market | 0 | 2 | 8 | 0 | 0 | 3 | 6 | 1 |

*(continued)*

| Outcome Type | 2012–16 (*N* = 119) | | | | 2017–20 (*N* = 117) | | | |
|---|---|---|---|---|---|---|---|---|
| | CDF | FIG | INR | MAS | CDF | FIG | INR | MAS |
| 11. Integration in the market | 9 | 2 | 1 | 3 | 0 | 3 | 3 | 4 |
| 12. Inclusiveness in the market | 0 | 7 | 0 | 1 | 0 | 5 | 0 | 0 |
| 13. Sustainability in the market | 1 | 0 | 1 | 4 | 0 | 1 | 1 | 3 |

*Source:* Independent Evaluation Group.

*Note:* CDF = Disruptive Technologies and Funds; GHG = greenhouse gas; FIG = Financial Institutions Group; IFC = International Finance Corporation; INR = Infrastructure; MAS = Manufacturing, Agribusiness, and Services; MSME = micro, small, and medium enterprise.

Projects are less likely to achieve market-level outcomes, particularly competitiveness in the market, than project-level outcome types (see table C.3). The assessment uses two metrics to assess the relationship between IFC's outcome types and the project performance of IFC projects, including (i) the achievement of outcome claims, referred to as outcome claim achieved, at the time of evaluation, and (ii) the change of AIMM ratings between the project's approval and monitoring stages. The competitiveness, integration, and sustainability market-level outcome types have the lowest achievement rates of all outcome types, at 36 percent, 43 percent, and 38 percent, respectively. Among the project-level outcome claims, access to goods and services had a relatively low success rate, although access to services for women had the highest achievement rate (71 percent).

IFC's achievement of market-level claims is less certain than its achievement of project-level claims, so market-level claims are considered riskier. Risk is measured by the changes in outcome claim ratings from the project's approval stage until its monitoring stage. Overall, market-level outcome types have a larger share of downgraded AIMM claim ratings (−1 or more) than project-level outcome types. Meanwhile, access to goods and services shows relatively large variance in the claim rating change among project-level outcome claims.

These results show that it is more difficult for IFC to achieve and measure market-level outcomes than project-level outcomes in general. This is because the success of market outcomes depends on the broader market environment and actions by external factors, such as government officials or private companies. Also, measuring market-level outcomes is challenging because of the long-term time horizons for outcomes to materialize, the challenge of attributing market-level results to IFC-supported projects, and the minimal impact that an individual IFC project can have on the broader market. It should be noted that in the AIMM framework, IFC's assertions explicitly focus on IFC contribution to market changes, rather than attribution. The Independent Evaluation Group's 2019 evaluation *Creating Markets to Leverage the Private Sector for Sustainable Development and Growth* reinforced this view by emphasizing the critical role an enabling environment plays in creating markets and calling for strengthened monitoring and evaluation systems for market creation projects (World Bank 2019). By contrast, the results shown in table C.3[4] suggest that project-level outcomes are less challenging to achieve and less risky. This is because these outcome types rely on providing goods, services, financing, or infrastructure, all of which IFC and its counterparts have more control over achieving. That is not to say that project-level outcomes are not challenging in their own way, only that they are easier to measure and depend much less on external factors. For example, "access to goods and services for micro, small, and medium enterprises" has a relatively lower achievement rate and larger variance in change ratings. This outcome type requires expanded lending to micro, small, and medium enterprises, enabling them to borrow from financial institutions, which is not entirely within the control of the project either. IFC is now striving to focus more on the use of goods and services within access outcome types.

## Table C.3. Outcome Type Performance and Performance Changes from the Approval Stage to the Monitoring Stage, 2012–20

| Claim Type | Claim Achieved (%) | Outcome Claims (no.) | Change in Claim Rating (%) −2 | −1 | 0 | +1 | +2 |
|---|---|---|---|---|---|---|---|
| Project-level claim | | | | | | | |
| 1.1 Access to goods and services (1.1.1–1.1.3) | 51 | 193 | 4 | 22 | 63 | 10 | 1 |
| 1.1.1 Access to goods and services (MSME) | 51 | 97 | 5 | 27 | 57 | 11 | 0 |
| 1.1.2 Access to goods and services (female) | 71 | 14 | 8 | 8 | 62 | 15 | 8 |
| 1.1.3 Access to goods and services (customers) | 53 | 97 | 2 | 15 | 73 | 10 | 0 |
| 1.2 Quality/affordability of goods and services | 63 | 16 | 13 | 13 | 67 | 7 | 0 |
| 2.1 Suppliers/distributors reached | 68 | 25 | 0 | 18 | 73 | 9 | 0 |
| 2.3 Improved sales/profitability of suppliers/distributors | 66 | 29 | 0 | 16 | 76 | 8 | 0 |
| 3.1 Increased employment | 57 | 56 | 2 | 16 | 73 | 9 | 0 |
| 6.2 GHG reduction | 70 | 30 | 0 | 23 | 77 | 0 | 0 |
| Market-level claim | | | | | | | |
| 9. Competitiveness in the market | 36 | 126 | 10 | 19 | 67 | 5 | 0 |
| 10. Resilience in the market | 63 | 24 | 4 | 26 | 70 | 0 | 0 |
| 11. Integration in the market | 43 | 28 | 11 | 11 | 75 | 4 | 0 |
| 12. Inclusiveness in the market | 69 | 16 | 6 | 38 | 44 | 13 | 0 |
| 13. Sustainability in the market | 38 | 13 | 0 | 46 | 54 | 0 | 0 |

*Source:* Independent Evaluation Group.

*Note:* (i) Of project-level outcome claims considered not achieved, 25 percent were considered not achieved because their results could not be verified. Of market-level outcome claims considered not achieved, 57 percent were considered not achieved because their results could not be verified. The relatively high percentage of nonverified market-level outcomes was partly due to the fact that many of those market outcomes were identified retroactively with backfilled AIMM. Monitoring of outcome results, particularly of market-level outcome, may improve under AIMM framework. (ii) "Change in Claim Rating" means step changes of intensity or movement ratings for claims from ex ante to monitoring. AIMM = Anticipated Impact Measurement and Monitoring; GHG = greenhouse gas; MSME = micro, small, and medium enterprise.

## Complete Tables for International Finance Corporation Outcome Type Analysis

### Table C.4. Outcome Types and Share in Total Projects, 2012–16, 2017–19, and Combined (percent)

| Outcome Type | 2012–16 (*N* = 119) | 2017–20 (*N* = 117) | 2012–20 (*N* = 236) |
|---|---|---|---|
| Project-level outcome | | | |
| 1.1 Access to goods and services (1.1.1–1.1.3) | 66 | 70 | 68 |
| 1.1.1 Access to goods and services (MSME) | 30 | 33 | 32 |
| 1.1.2 Access to goods and services (female) | 2 | 10 | 6 |
| 1.1.3 Access to goods and services (customers) | 39 | 42 | 40 |
| 1.2 Quality/affordability of goods and services | 8 | 6 | 7 |
| 1.3 Enhanced capacity of final beneficiaries | 18 | 9 | 13 |
| 1.4 Improved living standards (earnings) of individuals | 0 | 1 | 0 |
| 1.5 Improved sales/profitability of enterprises | 22 | 12 | 17 |
| 2.1 Suppliers/distributors reached | 6 | 15 | 10 |
| 2.2 Improved capacity of suppliers/distributors | 0 | 5 | 3 |
| 2.3 Improved sales/profitability of suppliers/distributors | 9 | 15 | 12 |
| 3.1 Increased employment | 19 | 28 | 24 |
| 3.2 Improved capacity/skills | 2 | 3 | 3 |
| 3.3 Improved earning of employees | 3 | 7 | 5 |
| 4.1 Increased transfers to the government | 4 | 3 | 4 |
| 5.1 Increased money spent/transfers to the communities | 0 | 0 | 0 |
| 6.1 Enhanced E&S standards of the client | 0 | 1 | 0 |
| 6.2 GHG reduction | 13 | 13 | 13 |
| 6.3 Efficient use of resources | 3 | 8 | 6 |
| 7.1 Gross value added | 3 | 3 | 3 |
| 7.2 Induced/indirect employment | 2 | 3 | 2 |

*(continued)*

| Outcome Type | 2012–16 (*N* = 119) | 2017–20 (*N* = 117) | 2012–20 (*N* = 236) |
|---|---|---|---|
| 7.3 Export sales | 0 | 1 | 0 |
| 8.1 Governance | 8 | 9 | 8 |
| Market-level outcome | | | |
| 9. Competitiveness in the market | 52 | 55 | 53 |
| 10. Resilience in the market | 9 | 11 | 10 |
| 11. Integration in the market | 14 | 9 | 12 |
| 12. Inclusiveness in the market | 8 | 5 | 6 |
| 13. Sustainability in the market | 6 | 5 | 6 |

*Source:* Independent Evaluation Group.

*Note:* E&S = environmental and social; GHG = greenhouse gas; MSME = micro, small, and medium enterprise.

## Table C.5. Outcome Types by Sector (percent)

| | 2012–16 (*N* = 119) | | | | 2017–20 (*N* = 117) | | | |
|---|---|---|---|---|---|---|---|---|
| Outcome Type | CDF | FIG | INR | MAS | CDF | FIG | INR | MAS |
| Project-level outcome | | | | | | | | |
| 1.1 Access to goods and services (1.1.1–1.1.3) | 8 | 30 | 18 | 9 | 3 | 37 | 17 | 13 |
| 1.1.1 Access to goods and services (MSME) | 8 | 22 | 0 | 0 | 3 | 28 | 0 | 3 |
| 1.1.2 Access to goods and services (female) | 0 | 2 | 0 | 0 | 0 | 10 | 0 | 0 |
| 1.1.3 Access to goods and services (customers) | 0 | 11 | 18 | 9 | 0 | 12 | 17 | 12 |
| 1.2 Quality/affordability of goods and services | 0 | 0 | 4 | 3 | 0 | 0 | 2 | 3 |
| 1.3 Enhanced capacity of final beneficiaries | 17 | 0 | 0 | 1 | 6 | 1 | 0 | 2 |
| 1.4 Improved living standards (earnings) of individuals | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 1.5 Improved sales/profitability of enterprises | 19 | 1 | 1 | 2 | 10 | 2 | 0 | 1 |
| 2.1 Suppliers/distributors reached | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 14 |

*(continued)*

| Outcome Type | 2012–16 (*N* = 119) | | | | 2017–20 (*N* = 117) | | | |
|---|---|---|---|---|---|---|---|---|
| | CDF | FIG | INR | MAS | CDF | FIG | INR | MAS |
| 2.2 Improved capacity of suppliers/distributors | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 |
| 2.3 Improved sales/profit-ability of suppliers/ distributors | 0 | 0 | 1 | 8 | 0 | 1 | 0 | 14 |
| 3.1 Increased employment | 0 | 0 | 8 | 11 | 0 | 0 | 7 | 22 |
| 3.2 Improved capacity/ skills | 0 | 0 | 0 | 2 | 0 | 0 | 1 | 3 |
| 3.3 Improved earning of employees | 0 | 0 | 0 | 3 | 0 | 0 | 1 | 6 |
| 4.1 Increased transfers to the government | 0 | 0 | 2 | 3 | 0 | 0 | 3 | 1 |
| 5.1 Increased money spent/transfers to the communities | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6.1 Enhanced E&S stan-dards of the client | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 6.2 GHG reduction | 0 | 0 | 10 | 3 | 1 | 0 | 9 | 3 |
| 6.3 Efficient use of resources | 0 | 1 | 2 | 1 | 1 | 4 | 0 | 3 |
| 7.1 Gross value added | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 3 |
| 7.2 Induced/indirect em-ployment | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 3 |
| 7.3 Export sales | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 8.1 Governance | 7 | 0 | 0 | 1 | 8 | 1 | 0 | 1 |
| Market-level outcome | | | | | | | | |
| 9. Competitiveness in the market | 9 | 20 | 13 | 11 | 9 | 23 | 12 | 11 |
| 10. Resilience in the market | 0 | 2 | 8 | 0 | 0 | 3 | 6 | 1 |
| 11. Integration in the market | 9 | 2 | 1 | 3 | 0 | 3 | 3 | 4 |
| 12. Inclusiveness in the market | 0 | 7 | 0 | 1 | 0 | 5 | 0 | 0 |
| 13. Sustainability in the market | 1 | 0 | 1 | 4 | 0 | 1 | 1 | 3 |

*Source:* Independent Evaluation Group.

*Note:* CDF = Disruptive Technologies and Funds; E&S = environmental and social; FIG = Financial Institutions Group; INR = Infrastructure; GHG = greenhouse gas; MAS = Manufacturing, Agribusiness, and Services; MSME = micro, small, and medium enterprise.

**Table C.6.** Average Claim Achieved and Claim Rating Variance

| Claim Type | Claim Achieved (%) | Outcome Claims (no.) | Change in Claim Rating (%) | | | | |
|---|---|---|---|---|---|---|---|
| | | | −2 | −1 | 0 | +1 | +2 |
| Project-level claim | | | | | | | |
| 1.1 Access to goods and services (1.1.1–1.1.3) | 51 | 193 | 4 | 22 | 63 | 10 | 1 |
| 1.1.1 Access to goods and services (MSME) | 51 | 97 | 5 | 27 | 57 | 11 | 0 |
| 1.1.2 Access to goods and services (female) | 71 | 14 | 8 | 8 | 62 | 15 | 8 |
| 1.1.3 Access to goods and services (customers) | 53 | 97 | 2 | 15 | 73 | 10 | 0 |
| 1.2 Quality/affordability of goods and services | 63 | 16 | 13 | 13 | 67 | 7 | 0 |
| 1.3 Enhanced capacity of final beneficiaries | 45 | 31 | 0 | 9 | 91 | 0 | 0 |
| 1.4 Improved living standards (earnings) of individuals | 0 | 1 | 0 | 0 | 100 | 0 | 0 |
| 1.5 Improved sales/profitability of enterprises | 48 | 44 | 3 | 24 | 71 | 3 | 0 |
| 2.1 Suppliers/distributors reached | 68 | 25 | 0 | 18 | 73 | 9 | 0 |
| 2.2 Improved capacity of suppliers/distributors | 67 | 6 | 0 | 17 | 83 | 0 | 0 |
| 2.3 Improved sales/profitability of suppliers/distributors | 66 | 29 | 0 | 16 | 76 | 8 | 0 |
| 3.1 Increased employment | 57 | 56 | 2 | 16 | 73 | 9 | 0 |
| 3.2 Improved capacity/skills | 50 | 6 | 20 | 0 | 80 | 0 | 0 |
| 3.3 Improved earning of employees | 58 | 12 | 0 | 27 | 64 | 9 | 0 |
| 4.1 Increased transfers to the government | 44 | 9 | 0 | 11 | 89 | 0 | 0 |
| 5.1 Increased money spent/transfers to the communities | — | 0 | — | — | — | — | — |
| 6.1 Enhanced E&S standards of the client | 0 | 1 | 0 | 0 | 100 | 0 | 0 |
| 6.2 GHG reduction | 70 | 30 | 0 | 23 | 77 | 0 | 0 |
| 6.3 Efficient use of resources | 57 | 14 | 0 | 15 | 85 | 0 | 0 |
| 7.1 Gross value added | 83 | 6 | 0 | 17 | 83 | 0 | 0 |

*(continued)*

| Claim Type | Claim Achieved (%) | Outcome Claims (no.) | Change in Claim Rating (%) | | | | |
|---|---|---|---|---|---|---|---|
| | | | −2 | −1 | 0 | +1 | +2 |
| 7.2 Induced/indirect employment | 40 | 5 | 0 | 67 | 33 | 0 | 0 |
| 7.3 Export sales | 100 | 1 | 0 | 0 | 100 | 0 | 0 |
| 8.1 Governance | 55 | 20 | 0 | 17 | 83 | 0 | 0 |
| Market-level claim | | | | | | | |
| 9. Competitiveness in the market | 36 | 126 | 10 | 19 | 67 | 5 | 0 |
| 10. Resilience in the market | 63 | 24 | 4 | 26 | 70 | 0 | 0 |
| 11. Integration in the market | 43 | 28 | 11 | 11 | 75 | 4 | 0 |
| 12. Inclusiveness in the market | 69 | 16 | 6 | 38 | 44 | 13 | 0 |
| 13. Sustainability in the market | 38 | 13 | 0 | 46 | 54 | 0 | 0 |

*Source:* Independent Evaluation Group.

*Note:* E&S = environmental and social; GHG = greenhouse gas; MSME = micro, small, and medium enterprise.

## Table C.7. Outcome Typologies for IFC Investment Projects

| Outcome Type (Name) | Description |
|---|---|
| 1.1.1 Access to services (MSMEs) | Increased number of MSMEs as final beneficiaries of goods and services of the project or company. Increased volume of goods and services produced or provided by the project or company can be considered under this outcome type. |
| 1.1.2 Access to services (female) | Increased number of final female beneficiaries of goods and services of the project or company. |
| 1.1.3 Access to services (customers) | Increased number of individual customers as final beneficiaries of goods and services of the project or company. Customers of utility services are representative of this group. Increased volume of goods and services produced or provided by the project or company can be considered under this outcome type. |
| 1.2 Quality and affordability of goods and services | Improved quality of goods and services produced by the project or company, compared with the baseline or with other producers or providers. Lower production costs or process are included.<br>Reduced prices of goods and services, compared with the baseline or other produces or providers, are also included here. |

(continued)

| Outcome Type (Name) | Description |
|---|---|
| 1.3 Enhanced capacity of final beneficiaries | Enhanced capacity of the final beneficiaries as a result of advisory services or training that is part of the project scope. |
| 1.4 Improved living standards (earnings) of individuals | Increased revenue or reduced expenditure by the final beneficiaries (individuals) of goods and services produced by the project or company. |
| 1.5 Improved sales and profitability of enterprises | Increased revenue or reduced expenditure or overall productivity by the final beneficiaries (enterprises) of goods and services produced by the project or company. |
| 2.1 Suppliers and distributors reached | Increased number of suppliers who provide inputs to the project or company, or the project expands the network of distributors of goods or services produced by the project or company. |
| 2.2 Improved capacity of suppliers and distributors | Capacity of suppliers or distributors improved as a result of advisory services or training that is part of the project scope. |
| 2.3 Improved sales and profitability of suppliers and distributors | The project increases the volume of inputs provided by its suppliers, or the project increases the goods or services to be distributed by its distributors. |
| 3.1 Increased employment | Increased direct employment of the client company. |
| 3.2 Improved capacity or skills | Training is provided to the employees of the project or company. |
| 3.3 Improved earning of employees | Increased wages to employees of the project or company. |
| 4.1 Increased transfers to the government | Payment by projects or companies to the government, such as in the form of taxes, royalties, fees, or dividends. |
| 5.1 Increased money spent or transfer to the communities | Payment to the communities in relation to the project or company, such as health, educational, or vocational programs in association with infrastructure projects. |
| 6.1 Enhanced E&S standards of the client | IFC supports its clients to enhance their E&S standards. |
| 6.2 GHG reduction | Projects such as renewable energy or energy efficiency projects contribute to the reduction or avoidance of GHGs. |
| 6.3 Efficient use of resources | The project will reduce the use of water and other resources, or the project promotes solid waste management and implements a waste-to-energy project. |
| 7.1 Gross value added | The project brings gross value added to the economy, which is calculated based on a multiplier and expressed in monetary value. |
| 7.2 Induced or indirect employment | Induced and indirect employment results from the project. This is also based on the multipliers. |
| 7.3 Export sales | The project increases export of goods and services produced. The economy's external balance from the generation and consumption of foreign currency. |
| 8.1 Governance | Enhanced governance or capacity of IFC's client company. |

*(continued)*

| Outcome Type (Name) | Description |
|---|---|
| 9 Competitiveness in the market | Competitive markets are those where firms can effectively enter, exit, and compete, and in which they can innovate and strive for efficiency under fair and good regulatory governance. Specific elements include the following:<br>» Market structure and functioning;<br>» Change in price;<br>» New practices, technology, product innovation (first movers);<br>» Product and business model differentiation, change in product offering, or greater value addition; and<br>» Enhanced efficiency under fair and good regulatory governance (including accreditation). |
| 10 Resilience in the market | Making markets more resilient involves improving the depth, structure, regulation, and governance of markets to help them withstand physical, financial, or economic shocks and stresses. Climate resilience, in specific sectors, is also important in helping markets withstand climate-related shocks and stresses. Resilient markets support growth without excessive volatility and destabilizing economic reversals. Specific elements include the following:<br>» Improved corporate governance of the direct clients;<br>» Diversification (for example, energy sources or funding sources in sectors or products);<br>» Capacity to face shocks and stress;<br>» Improved depth, structure, regulation, and governance of market (capacity of institutional body to regulate the sector);<br>» Effect on domestic supply volatility and energy security; and<br>» Financial stability and consumer protection. |
| 11 Integration in the market | Enhancing physical or financial connectivity, within and across markets, to support greater market integration. Examples include stronger integration with financial markets and growing domestic and global value chains to pave the way for products or structures that mobilize resources at scale. Specific elements include the following:<br>» Enhanced physical or financial connectivity;<br>» Geographical or spatial integration;<br>» Integration with financial markets (including capital mobilization);<br>» Data integration;<br>» Growing domestic and global value chains;<br>» Trade diversification; and<br>» Economic complexity. |

*(continued)*

| Outcome Type (Name) | Description |
|---|---|
| 12 Inclusiveness in the market | Inclusive markets support fair and full access to all to goods and services, as well as to finance and economic opportunities. Increasing inclusiveness includes improving access for underserviced groups. Specific elements include the following:<br>» Marketwide focus and access for underserviced groups;<br>» Marketwide enabling framework or standards supporting inclusive business; and<br>» Enhanced diversity. |
| 13 Sustainability in the market | When firms and consumers adopt climate-related, environmentally and socially sustainable products, technologies, and practices, they promote greater market sustainability. This is key to helping firms and industries apply environmentally and socially sustainable approaches to mitigate risk, realize opportunities, and maximize operational efficiency. Specific elements include the following:<br>» Climate-related, environmentally and socially sustainable products, technologies, standards, and practices adopted;<br>» Conducive legal or regulatory framework to foster sustainability; and<br>» Broad capacity and supporting institutions or sustainability practice. |

*Source:* Independent Evaluation Group.

*Note:* E&S = environmental and social; GHG = greenhouse gas; IFC = International Finance Corporation; MSME = micro, small, and medium enterprise.

## Multilateral Investment Guarantee Agency Outcome Type Analysis

The Multilateral Investment Guarantee Agency's (MIGA) outcome types reflect its focus on larger infrastructure projects with potential foreign investment promotion effect. MIGA has been pursuing outcomes that are directly derived from its projects as project-level outcomes and outcomes that have broader impact to promote foreign investment. Among MIGA projects, access to goods and services for customers and market development have been the most common outcome types for project outcome level and foreign investment–level outcome, accounting for 70 percent and 47 percent of projects, respectively, for the combined fiscal year (FY)12–14 and FY17–19 periods (table C.8). These reflect MIGA's focus on large infrastructure projects with broader effects to promote foreign investment. In addition, it has been increasingly supporting projects with demonstration effects, which is reflected in the increasing share of market outcome types such as business and sector practices, and signaling effects.

MIGA projects have a higher probability of achieving project-level outcomes than foreign investment–level outcomes. For example, table C.9 shows that access to goods and services for customers, quality and affordability of goods and services, and increased employment have the highest probabilities of success. From FY12–14 to FY17–19, the achievement rates increased substantially for project-level outcomes such as access for micro, small, and medium enterprises and access for customers. Meanwhile, market development achievement rates decreased. MIGA's inherent limitations as a guarantee provider in collecting data on the development results of projects means that many projects lack sufficient evidence to rate project outcomes. This suggests that MIGA's improved development outcome ratings are due to both increased evidence collection in recent years and actual improvement in performance.

## Table C.8. Share of MIGA Projects with Certain Outcome Types (percent)

| Outcome Type | 2012–14 (*N* = 26) | 2017–19 (*N* = 39) |
|---|---|---|
| Project-level claim | | |
| 1.1 Access (1.1.1–1.1.3) | 65 | 85 |
| 1.1.1 Access to goods and services (SME) | 27 | 18 |
| 1.1.2 Access to goods and services (female) | 8 | 0 |
| 1.1.3 Access to goods and services (other) | 58 | 77 |
| 1.2 Quality/affordability of goods and services | 38 | 51 |
| 1.3 Enhanced capacity of final beneficiaries | 35 | 5 |
| 1.6 Economic return | 50 | 23 |
| 1.7 Financial/business performance of direct clients | 38 | 38 |
| 3.1 Increased employment | 35 | 54 |
| 3.2 Improved capacity/skills | 38 | 28 |
| 4.1 Increased transfers to the government | 38 | 44 |
| 8.1 Governance | 35 | 8 |
| Foreign investment–level claim | | |
| 9. Business and sector practices | 19 | 33 |
| 10. Market development | 62 | 41 |
| 11. Development reach | 0 | 0 |
| 12. Sustainability | 0 | 0 |
| 13. Signaling effects | 0 | 3 |

*Source:* Independent Evaluation Group.

*Note:* Individual projects can have multiple outcome types. MIGA = Multilateral Investment Guarantee Agency; SME = small and medium enterprise.

**Table C.9.** Outcome Type Performance from the Approval Stage to the Evaluation Stage, 2012–14 and 2017–19

| Outcome Type Performance | 2012–14 Outcome Claims (% achieved) | (no.) | 2017–19 Outcome Claims (% achieved) | (no.) |
|---|---|---|---|---|
| Project-level outcome | | | | |
| 1.1 Access (1.1.1–1.1.3) | 33 | 27 | 56 | 54 |
| 1.1.1 Access to goods and services (MSME) | 11 | 9 | 56 | 9 |
| 1.1.2 Access to goods and services (female) | 0 | 2 | | 0 |
| 1.1.3 Access to goods and services (customers) | 42 | 19 | 57 | 46 |
| 1.2 Quality/affordability of goods and services | 46 | 13 | 52 | 29 |
| 1.3 Enhanced capacity of final beneficiaries | 33 | 9 | 50 | 2 |
| 1.4 Improved living standards (earnings) of individuals | — | 0 | 100 | 2 |
| 2.1 Suppliers/distributors reached | 0 | 2 | 100 | 2 |
| 2.3 Improved sales/profitability of suppliers/distributors | 0 | 5 | 33 | 9 |
| 3.1 Increased employment | 54 | 13 | 38 | 21 |
| 4.1 Increased transfers to the government | 30 | 10 | 33 | 18 |
| 6.2 GHG reduction | — | 0 | 57 | 7 |
| 6.3 Efficient use of resources | 40 | 5 | 71 | 7 |
| Foreign-level investment | | | | |
| 9. Business and sector practices | 40 | 5 | 40 | 15 |
| 10. Market development | 39 | 23 | 29 | 24 |
| 13. Signaling effects | — | 0 | 100 | 1 |

*Source:* Independent Evaluation Group.

*Note:* Outcome type estimation by Independent Evaluation Group based on the approval documents and confirmed cases at the evaluation. Of the outcome claims considered not achieved, 51 percent were considered not achieved because their results could not be verified. Broken down by outcome types, foreign investment–level outcome claims had a higher share of unverified claims (61 percent) than project-level outcome claims (50 percent). The level is relatively high, particularly for foreign investment–level outcome claims, and it would be expected that the tracking of outcome claim results is strengthened under the newly introduced Impact Measurement and Project Assessment Comparison Tool framework. — = not available; GHG = greenhouse gas; MSME = micro, small, and medium enterprise.

# Additional Tables for MIGA Outcome Type Analysis

## Table C.10. MIGA Outcome Types by Project Share (percent)

| Outcome Type | 2012–14 (*N* = 26) | 2017–19 (*N* = 39) |
|---|---|---|
| Project-level claim | | |
| 1.1 Access (1.1.1–1.1.3) | 65 | 85 |
| 1.1.1 Access to goods and services (SME) | 27 | 18 |
| 1.1.2 Access to goods and services (female) | 8 | 0 |
| 1.1.3 Access to goods and services (other) | 58 | 77 |
| 1.2 Quality/affordability of goods and services | 38 | 51 |
| 1.3 Enhanced capacity of final beneficiaries | 35 | 5 |
| 1.4 Improved living standards (earnings) of individuals | 0 | 5 |
| 1.5 Profitability of direct clients | 8 | 10 |
| 1.6 Economic return | 50 | 23 |
| 1.7 Financial/business performance of direct clients | 38 | 38 |
| 2.1 Suppliers/distributors reached | 8 | 5 |
| 2.2 Improved capacity of suppliers/distributors | 0 | 5 |
| 2.3 Improved sales/profitability of suppliers/distributors | 12 | 23 |
| 3.1 Increased employment | 35 | 54 |
| 3.2 Improved capacity/skills | 38 | 28 |
| 3.3 Improved earning of employees | 19 | 10 |
| 4.1 Increased transfers to the government | 38 | 44 |
| 5.1 Increased money spent/transfers to the communities | 0 | 5 |
| 6.1 Enhanced E&S standards of the client | 0 | 5 |
| 6.2 GHG reduction | 0 | 18 |
| 6.3 Efficient use of resources | 15 | 13 |
| 7.1 Gross value added | 19 | 28 |
| 7.2 Induced/indirect employment | 8 | 13 |
| 7.3 Export sales | 8 | 8 |
| 8.1 Governance | 35 | 8 |
| Foreign investment–level claim | | |
| 9. Business and sector practices | 19 | 33 |
| 10. Market development | 62 | 41 |
| 11. Development reach | 0 | 0 |
| 12. Sustainability | 0 | 0 |
| 13. Signaling effects | 0 | 3 |

*Source:* Independent Evaluation Group.

(continued)

*Note:* E&S = environmental and social; GHG = greenhouse gas; MIGA = Multilateral Investment Guarantee Agency; SME = small and medium enterprise.

## Table C.11. MIGA Outcome Claim Achieved per Period

| Claim Type | 2012–14 Claim Achieved (%) | 2012–14 Outcome Claims (no.) | 2017–19 Claim Achieved (%) | 2017–19 Outcome Claims (no.) |
|---|---|---|---|---|
| Project-level claim | | | | |
| 1.1 Access (1.1.1–1.1.3) | 33 | 27 | 56 | 54 |
| 1.1.1 Access to goods and services (SME) | 11 | 9 | 56 | 9 |
| 1.1.2 Access to goods and services (female) | 0 | 2 | — | 0 |
| 1.1.3 Access to goods and services (other) | 42 | 19 | 57 | 46 |
| 1.2 Quality/affordability of goods and services | 46 | 13 | 52 | 29 |
| 1.3 Enhanced capacity of final beneficiaries | 33 | 9 | 50 | 2 |
| 1.4 Improved living standards (earnings) of individuals | — | 0 | 100 | 2 |
| 1.5 Profitability of direct clients | 33 | 3 | 25 | 4 |
| 1.6 Economic return | 31 | 13 | 60 | 10 |
| 1.7 Financial/business performance of direct clients | 40 | 10 | 40 | 15 |
| 2.1 Suppliers/distributors reached | 0 | 2 | 100 | 2 |
| 2.2 Improved capacity of suppliers/distributors | — | 0 | 0 | 4 |
| 2.3 Improved sales/profitability of suppliers/distributors | 0 | 5 | 33 | 9 |
| 3.1 Increased employment | 54 | 13 | 38 | 21 |
| 3.2 Improved capacity/skills | 45 | 11 | 33 | 12 |
| 3.3 Improved earning of employees | 80 | 5 | 25 | 4 |
| 4.1 Increased transfers to the government | 30 | 10 | 33 | 18 |
| 5.1 Increased money spent/transfers to the communities | — | 0 | 0 | 2 |
| 6.1 Enhanced E&S standards of the client | — | 0 | 100 | 2 |
| 6.2 GHG reduction | — | 0 | 57 | 7 |

*(continued)*

| Claim Type | 2012–14 Claim Achieved (%) | 2012–14 Outcome Claims (no.) | 2017–19 Claim Achieved (%) | 2017–19 Outcome Claims (no.) |
|---|---|---|---|---|
| 6.3 Efficient use of resources | 40 | 5 | 71 | 7 |
| 7.1 Gross value added | 33 | 6 | 55 | 11 |
| 7.2 Induced/indirect employment | 0 | 2 | 50 | 6 |
| 7.3 Export sales | 0 | 2 | 100 | 4 |
| 8.1 Governance | 44 | 9 | 0 | 4 |
| Foreign investment–level claim | | | | |
| 9. Business and sector practices | 40 | 5 | 40 | 15 |
| 10. Market development | 39 | 23 | 29 | 24 |
| 11. Development reach | — | 0 | — | 0 |
| 12. Sustainability | — | 0 | — | 0 |
| 13. Signaling effects | — | 0 | 100 | 1 |

*Source:* Independent Evaluation Group.

*Note:* E&S = environmental and social; GHG = greenhouse gas; MIGA = Multilateral Investment Guarantee Agency; SME = small and medium enterprise.

## Table C.12. Outcome Typologies for MIGA Projects

| Outcome Type (Name) | Description |
|---|---|
| 1.1.1 Access to services (MSMEs) | Increased number of MSMEs as final beneficiaries of goods and services of the project or company. Increased volume of goods and services produced or provided by the project or company can be considered under this outcome type. |
| 1.1.2 Access to services (female) | Increased number of final female beneficiaries of goods and services of the project or company. |
| 1.1.3 Access to services (customers) | Increased number of individual customers as final beneficiaries of goods and services of the project or company. Customers of utility services are representative of this group. Increased volume of goods and services produced or provided by the project or company can be considered under this outcome type. |

*(continued)*

| Outcome Type (Name) | Description |
|---|---|
| 1.2 Quality and affordability of goods and services | Improved quality of goods and services produced by the project or company, compared with the baseline or with other producers or providers. Lower production costs or process are included.<br><br>Reduced prices of goods and services, compared with the baselines or other producers or providers, are also included here. |
| 1.3 Enhanced capacity of final beneficiaries | Enhanced capacity of the final beneficiaries as a result of advisory services or training that is part of the project scope. |
| 1.4 Improved living standards (earnings) of individuals | Increased revenue or reduced expenditure by the final beneficiaries (individuals) of goods and services produced by the project or company. |
| 1.5 Improved sales or profitability of enterprises | Increased revenue, reduced expenditure, or increased overall productivity by the final beneficiaries (enterprises) of goods and services produced by the project or company. |
| 1.6 Economic return | Economic rate of return. |
| 1.7 Financial and business performance of direct clients | Financial and business performance of direct clients, mostly project-executing agencies. |
| 2.1 Suppliers and distributors reached | Increased number of suppliers who provide inputs to the project or company, or the project expands the network of distributors of goods or services produced by the project or company. |
| 2.2 Improved capacity of suppliers and distributors | Capacity of suppliers or distributors improved as a result of advisory services or training that is part of the project scope. |
| 2.3 Improved sales or profitability of suppliers and distributors | The project increases the volume of inputs provided by its suppliers or increases the goods or services to be distributed by its distributors. |
| 3.1 Increased employment | Increased direct employment of the client company. |
| 3.2 Improved capacity or skills | Training is provided to the employees of the project or company. |
| 3.3 Improved earning of employees | Increased wages to employees of the project or company. |
| 4.1 Increased transfers to the government | Payments from the projects or companies to the government, such as taxes, royalties, fees, or dividends. |
| 5.1 Increased money spent or transferred to the communities | Payment to the communities in relation to the project or company, such as health, educational, or vocational programs in association with infrastructure projects. |
| 6.1 Enhanced E&S standards of the client | IFC supports its clients to enhance their E&S standards. |
| 6.2 GHG reduction | Projects such as renewable energy projects and energy efficiency projects that contribute to the reduction or avoidance of greenhouse gases. |

| Outcome Type (Name) | Description |
|---|---|
| 6.3 Efficient use of resources | The project will reduce use of water and other resources, or the project promotes solid waste management or implements a waste-to-energy project. |
| 7.1 Gross value added | The project brings gross value added to the economy, which is calculated based on a multiplier and expressed in monetary value. |
| 7.2 Induced or indirect employment | Induced and indirect employment as a result of the project. This is also based on the multipliers. |
| 7.3 Export sales | The project increases export of goods and services produced. The economy's external balance from the generation and consumption of foreign currency. |
| 8.1 Governance | Enhanced governance or capacity of MIGA's client company. |
| 9 Business and sector practices | Potential to improve (financial or operational) performance of future investments through demonstration or transfer of new technologies, capabilities, practices, or business models. |
| 10 Market development | Potential to enhance the market structure, potentially also benefiting future investors (for example, competitiveness, resilience, supply chain integration, regulatory environment, and so on). |
| 11 Development reach | Potential to inspire future investments to focus more on inclusiveness (for example, reaching underserviced populations such as minorities, women, and so on). |
| 12 Sustainability | Potential to inspire future investments to improve E&S sustainability through adoption of better standards, practices, and so on. |
| 13 Signaling effects | Potential to inspire or encourage further foreign investment (where there may have been some real or perceived barriers). |

*Source:* Independent Evaluation Group.

*Note:* E&S = environmental and social; GHG = greenhouse gas; IFC = International Finance Corporation; MIGA = Multilateral Investment Guarantee Agency; MSME = micro, small, and medium enterprise.

## Reference

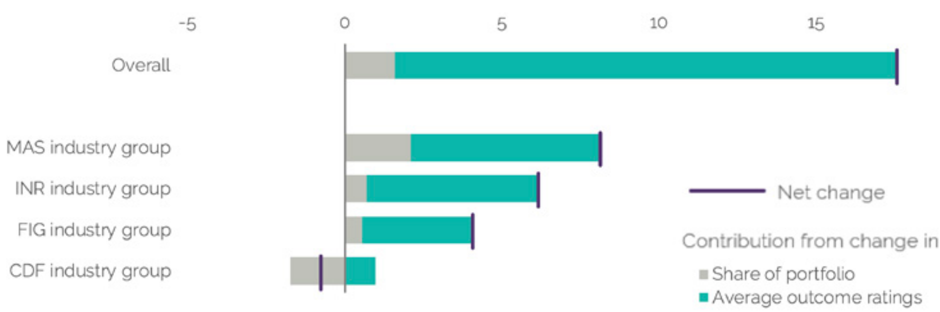
World Bank. 2019. *Creating Markets to Leverage the Private Sector for Sustainable Development and Growth*. Independent Evaluation Group. Washington, DC: World Bank. https://ieg.worldbankgroup.org/evaluations/creating-markets.

# Appendix D. International Finance Corporation Decomposition Analysis

## Decomposition Analysis for International Finance Corporation Projects

The decomposition analysis for International Finance Corporation project ratings between 2018 and 2019 shows that there were certain changes in the portfolio composition of the project categories, and overall rating improvements made major contributions across the categories (figures D.1, D.2, D.3, D.4, and D.5).

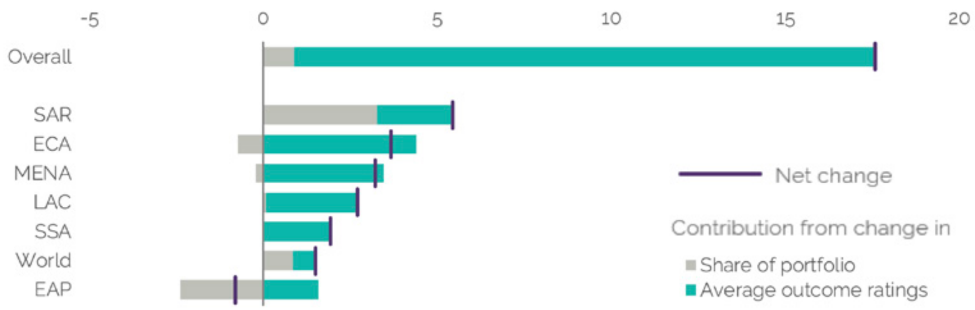
**Figure D.1.** Industry Group




*Source:* Independent Evaluation Group.

*Note:* CDF = Disruptive Technologies and Funds; FIG = Financial Institutions Group; INR = Infrastructure; MAS = Manufacturing, Agribusiness, and Services.

## Figure D.2. Region



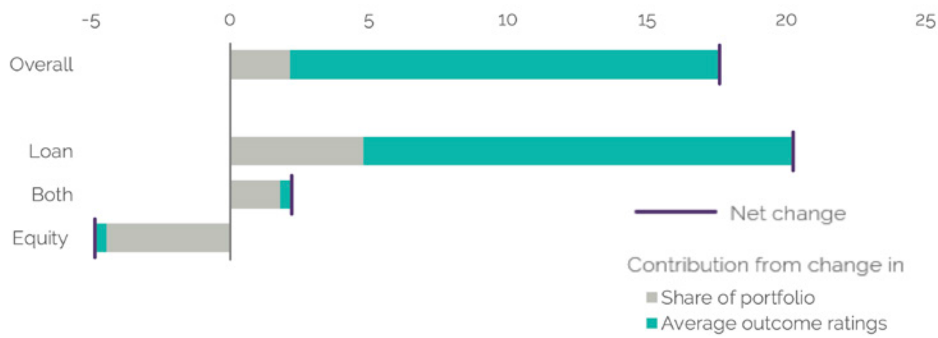

*Source:* Independent Evaluation Group.

*Note:* EAP = East Asia and Pacific; ECA = Europe and Central Asia; LAC = Latin America and the Caribbean; MENA = Middle East and North Africa; SAR = South Asia; SSA = Sub-Saharan Africa.

## Figure D.3. Investment Size


*Source:* Independent Evaluation Group.

## Figure D.4. Instrument Type




*Source:* Independent Evaluation Group.

## Figure D.5. Subsectors


*Source:* Independent Evaluation Group.

*Note:* CDF = Disruptive Technologies and Funds; FIG = Financial Institutions Group, MAS = Manufacturing, Agribusiness, and Services.

IEG Independent Evaluation Group
WORLD BANK GROUP
World Bank • IFC • MIGA
WHAT WORKS

The World Bank
1818 H Street NW
Washington, DC 20433