Results and Performance of the World Bank Group 2023

An Independent Evaluation

IEG
Independent Evaluation Group

WORLD BANK GROUP
World Bank • IFC • MIGA

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

ATTRIBUTION
Please cite the report as: World Bank. 2023. *Results and Performance of the World Bank Group 2023*. Independent Evaluation Group. Washington, DC: World Bank.

COVER PHOTO
Adobe Stock/chris3d

EDITING AND PRODUCTION
Amanda O'Brien

GRAPHIC DESIGN
Luísa Ulhoa
Rafaela Sarinho

Results and Performance of the World Bank Group 2023

An Independent Evaluation

December 7, 2023

Contents

Boxes

Figures

Tables

Appendixes

Abbreviations

AIMM	Anticipated Impact Measurement and Monitoring
CY	calendar year
DPF	development policy financing
FCS	fragile and conflict-affected situation
FY	fiscal year
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
IMPACT	Impact Measurement and Project Assessment Comparison Tool
IPF	investment project financing
M&E	monitoring and evaluation
MIGA	Multilateral Investment Guarantee Agency
PBP	project business performance
PDO	project development objective
PforR	Program-for-Results
PPP	public-private partnership
RAP	*Results and Performance of the World Bank Group*
XPSR	Expanded Project Supervision Report

All dollar amounts are US dollars unless otherwise indicated.

Acknowledgments

This report was prepared by an Independent Evaluation Group team led by Mercedes Vellez (senior evaluation officer) and Unurjargal Demberel (evaluation officer), under the overall direction of Sabine Bernabè (Director-General, Evaluation) and the supervision of Carmen Nonay (director, Finance, Private Sector, Infrastructure, and Sustainable Development, and acting director, Human Development and Economic Management), Theo David Thomas (director, Human Development and Economic Management), Galina Sotirova (manager, Corporate and Human Development Unit), and Beata Lenard (manager, Financial and Private Sector Micro Unit).

The core team for the analysis of the World Bank's portfolio included Chikako Miwa, Gaby Loibl, Kavita Mathur, Virginia Ziulu, and Xiaoxiao Peng. Other Independent Evaluation Group colleagues and consultants who made valuable contributions were Andrei Wong Espejo, Ariya Hagh, Orlando Vaca Rodriguez, Rocio Manchado Garabito, Santiago Ramirez, and Shiva Sharma. For the analysis of the International Finance Corporation's and the Multilateral Investment Guarantee Agency's portfolios, the core team members included Carlos Stagliano, Fan Yang, and Min Kyu Song. Other Financial and Private Sector Micro Unit colleagues who provided valuable input were Ananda Ghose, Belen Barbeito, Feruza Abduazimova, Ichiro Toda, Izlem Yenice, Leonardo Bravo, and Sunbol Waheed. Maximillian Ashwill was the report's primary editor.

The external advisory panel included Tamar Manuelyan Atinc (nonresident senior fellow at the Brookings Institution), Marie Gaarder (executive director, International Initiative for Impact Evaluation), Lavagnon Ika (professor, University of Ottawa), and Roland Michelitsch (independent international development consultant).

The team is grateful to Rasmus Heltberg and Andrew Stone for their comments on an early version and to other Independent Evaluation Group colleagues who shared their insight and feedback throughout the evaluation. The team also thanks Operations Policy and Country Services, International Finance Corporation, and Multilateral Investment Guarantee Agency colleagues who engaged with us during the preparation of the report.

Overview

This is the 13th annual *Results and Performance of the World Bank Group (RAP)* report. The *RAP* comprehensively reviews evidence from the evaluation and validation work of the Independent Evaluation Group (IEG). It assesses the development effectiveness of the World Bank Group, including the World Bank, comprising the International Bank for Reconstruction and Development, the International Development Association (IDA), the International Finance Corporation (IFC), and the Multilateral Investment Guarantee Agency (MIGA).

The analyses and methodologies presented in this *RAP* were guided by three principles: continuity, innovation, and symmetry. The report embraces continuity by building on previous *RAP*s, innovation by incorporating novel data and evaluation methods, and symmetry by conducting similar analyses across different Bank Group institutions. In particular, this *RAP* focuses on the evolution of project performance ratings across the three institutions. It also examines development outcomes underlying project performance ratings and assesses the validity of the monitoring and evaluation (M&E) frameworks tasked with measuring them. Furthermore, it investigates the relationship between the validity of indicators and project performance ratings. Lastly, this report analyzes the factors that affected the implementation and performance of Bank Group projects, particularly within the context of the pandemic. Since results and performance across Bank Group institutions are not comparable because of differing mandates, project restructuring arrangements, and project evaluation and rating methodologies, this *RAP*'s findings are presented separately for each institution (for more information on evaluation approaches, see chapter 1 and appendix A).

This is the first *RAP* with a substantial number of projects that were implemented during the COVID-19 pandemic. World Bank, IFC, and MIGA projects faced many obstacles during implementation, especially disruptions associated with the COVID-19 pandemic. The majority of Bank Group projects reported pandemic-related implementation challenges, although most of these projects began implementation well before the pandemic started. As a

share of project time length, Bank Group projects' exposure to the pandemic was rather limited, especially for the World Bank: on average, 14 percent of World Bank projects' lives occurred during the pandemic compared with 24 percent of IFC investment and 27 percent of MIGA guarantee projects' lives. As a result of the limited exposure of these projects to the pandemic and the presence of a sample selection bias in the *RAP* sample, this *RAP* does not fully capture the impact of COVID-19 on projects' performance. It is likely that project cohorts in the forthcoming years will exhibit a more accurate depiction of the far-reaching consequences of the COVID-19 pandemic.

World Bank

Project Performance

World Bank projects maintained or improved their performance ratings, as evaluated by IEG, between fiscal year (FY)21 and FY22 despite the effects of the COVID-19 pandemic and the Russian invasion of Ukraine. Both the average outcome and Bank performance ratings of the 181 investment project financing (IPF) and Program-for-Results projects in FY22 remained at 4.3 on a 6-point scale, as in FY20 and FY21—the highest average since FY12. IPF and Program-for-Results projects have seen an upward trajectory in M&E quality ratings, with the share of projects rated substantial or high increasing from 60 percent in FY21 to 63 percent in FY22. The average outcome rating for 17 development policy financing projects in FY22 also remained constant at an average of 4.0 on a 6-point scale, and the overall Bank performance ratings improved with an average rating increasing from 4.3 in FY21 to 4.6 in FY22 on a 6-point scale. Nevertheless, development policy financing performance ratings should be interpreted with caution because of the limited number of projects in the *RAP* sample (for more information on distribution of ratings, see chapter 2 and appendix B).

From a longer-term perspective, World Bank project performance ratings for IPF projects closing in FY20–22 also saw an improvement from previous years. This *RAP*'s in-depth analysis of 273 IPF projects revealed that projects exposed to the pandemic and closed during FY20–22 exhibited higher performance across all project ratings compared with projects that closed in FY18–20 and were not exposed to the pandemic. Efficacy ratings of intended

development outcomes have also shown consistent and statistically significant improvement from FY12–14 to FY20–22. This is particularly noteworthy as the size of the average International Bank for Reconstruction and Development and IDA net commitments has been increasing since FY12.

Explaining Project Performance

The limited exposure time of projects to the pandemic and the presence of a selection bias in the *RAP* sample help explain the World Bank's positive and healthy performance rating trends. All World Bank projects in the *RAP 2023* cohort were exposed to the pandemic to some extent during their project lives, and a vast majority reported being negatively affected by it. However, many projects were already at an advanced stage of implementation when the pandemic began, minimizing the severity of their exposure. In addition, the *RAP 2023* cohort may have an overrepresentation of successful projects with higher ratings because projects for which teams completed both an Implementation Completion and Results Report (ICR) and an Implementation Completion and Results Report Review (ICRR) relatively quickly after the project closure tended to have higher ratings than projects that delayed their ICRs and ICRRs. This pattern also applies to the rating trends and occurred in previous years, but it is even more likely to occur in this *RAP* cohort because of the evaluation cutoff date of December 2022. An exploration of the latest Implementation Status and Results Report ratings on progress toward achieving project development objectives in FY22 confirms this pattern. Projects with ICRRs completed and included in this *RAP* have higher average Implementation Status and Results Report ratings than projects with completed ICRs but in-progress ICRRs and even higher ratings than projects with uncompleted ICRs (see figure A.3). Therefore, the trends of project performance ratings should be interpreted carefully as they are likely to change downward in the future. This is especially true as more projects with extended exposure times to COVID-19 are evaluated and incorporated into the project rating trends (see appendix A for more details on the limitations of the data).

Despite the projects' limited exposure to it, the COVID-19 pandemic was the most salient challenge for World Bank project implementation during

FY20–22. According to the ICRs, COVID-19–related lockdowns and mobility restrictions led to economic downturns and disruptions in public services and institutional operations. Most projects reported implementation delays caused by supply chain shortages and other logistical challenges, which had an impact on civil works components of projects. The pandemic also led to the postponement of in-person project-related activities and, in some cases, the reallocation of project funds.

Other contextual, stakeholder, and project-related challenges not directly attributed to the COVID-19 crisis seemed to be more exacerbated in the *RAP 2023* cohort than in previous years. In particular, the low technical and organizational capacity of implementing agencies emerged as a key implementation constraint, especially in projects that failed to adequately identify and mitigate institutional capacity risks.

More project adaptation and timely restructuring also helped limit the impact of implementation challenges on project performance. Project restructurings, in the form of adjusted closing dates and results frameworks, were more common among FY20–22 projects than in previous years. These restructurings may have helped projects adapt to implementation challenges and respond to unexpected shocks, including those related to the pandemic. Furthermore, the timeliness of these course corrections mattered. Projects that made course corrections earlier in the project cycle had a higher likelihood of achieving their intended development outcomes than projects that adapted later, and, as a result, the early adapted projects received higher efficacy ratings.

The World Bank's improved M&E facilitated project adaptation and helped provide sufficient evidence on project achievements; however, some M&E outcome orientation challenges persist. M&E frameworks provide project teams with a deeper understanding of project achievements and challenges, enabling them to make well-informed adjustments during implementation. We confirm what previous *RAPs* suggested, namely, that World Bank projects with good-quality M&E tend to have higher efficacy ratings than projects with low-quality M&E. This *RAP*'s in-depth assessment of results framework indicators reinforced this positive influence of improved M&E quality on efficacy ratings. Nevertheless, some M&E outcome orientation challenges

persist, particularly in measuring institutional capacity-building outcomes. The attainment of these development outcomes still tends to be measured by intermediate outcome or lower-level indicators rather than outcome or higher-level indicators that demonstrate development impact, such as the improved capacity of public institutions to perform their functions.

International Finance Corporation

Project Performance

IFC investment project development outcome ratings declined only slightly despite projects' exposure to COVID-19 and the more challenging operating environment. The share of IFC's investment projects rated mostly successful or better, for development outcomes, decreased from 53 percent in calendar year (CY)19–21 to 50 percent in CY20–22, in line with IFC's Expanded Project Supervision Report self-ratings. IFC investment projects' exposure to COVID-19 averaged about 24 percent of their active project lives. The pandemic-related disruptions and economic slowdowns contributed to a particularly challenging operating environment for IFC's CY20–22 investment projects. That being said, IFC's *RAP 2023* cohort did not include investment projects that were severely affected by the pandemic but only those that were considered to have been moderately or minimally affected (as assessed by IFC at the time of sampling). It is also important to note that while IFC undertakes financial restructuring for its investment projects as needed, it has no formal procedures for modifying the original development objectives, indicators, and targets to adapt to changing market conditions.

IEG's ratings for IFC work quality, particularly for project preparation, declined in CY20–22. Overall, the share of IFC investment projects with satisfactory or better (high) IFC work quality ratings declined from 60 percent in CY19–21 to 55 percent in CY20–22. The share of investment projects with high project preparation work quality ratings decreased from 59 percent to 54 percent between CY19–21 and CY20–22. However, high supervision and administration work quality ratings stayed at approximately 70 percent during the same period. The *RAP 2023* confirmed the findings from previous *RAPs* that IFC work quality ratings for investment projects are positively and strongly associated with IFC's development outcome ratings.

IFC additionality success ratings in challenging environments were lower than the IFC average. Overall, the share of IFC investment projects with high additionality was 54 percent in CY20–22 (down from 59 percent in CY19–21). IFC additionality success ratings in African, fragile and conflict-affected situation (FCS), and IDA and blend countries were lower than the IFC average. Specifically, IFC realized its anticipated additionality in 37 percent of African projects, 33 percent of FCS projects, and 47 percent of IDA and blend investment projects. The gap between anticipated and realized additionality in these challenging markets was larger for nonfinancial additionality than for financial additionality. The *RAP 2023* confirmed the *RAP 2022* findings that IFC additionality ratings were positively and strongly correlated with project development outcome ratings.

Overall, IFC's investment outcome success ratings declined in CY20–22, although IFC's equity performance remained stable. IFC's overall investment outcome ratings have been satisfactory or better in 60 percent of investment projects in CY20–22, which was slightly lower than 64 percent in CY19–21. This decline was caused by the slight decline in loan investment outcome ratings. Among IFC investment projects with low loan investment outcome ratings, 42 percent had prepayments, which affected IFC's ability to realize full anticipated financial returns. In contrast, equity outcome ratings have remained stable, although only about one-third of equity investments generated satisfactory returns. A large share of IFC investment projects in challenging markets tended to not achieve IFC's "double bottom line" of delivering high development results and satisfactory investment returns. Development and investment outcome ratings were both low in 56 percent of FCS countries' investment projects, 51 percent of African projects, and 39 percent of IDA and blend projects, compared with an average of 31 percent for the IFC portfolio as a whole.

IFC advisory projects saw a slight performance decline in the more challenging operating environment. The development effectiveness success ratings of IFC advisory projects have been improving since FY15–17 but declined from 60 percent in FY19–21 to 54 percent in FY20–22. IFC's Project Completion Report self-ratings also showed a decline. External factors such as political conflicts, force majeure events, COVID-19–related disruptions, and client commitment issues also negatively affected the more recent projects in this

RAP cohort. Project design weaknesses and M&E shortcomings contributed to IFC advisory projects' low development effectiveness ratings. IFC's overall work quality remained satisfactory in 59 percent of advisory projects in FY20–22. However, IFC's preparation and design work quality ratings were satisfactory or better in fewer than half of projects in FY20–22. The implementation and supervision work quality success ratings of advisory projects declined marginally in FY20–22. The *RAP 2023* confirmed the findings from previous evaluations that development effectiveness ratings of advisory projects are highly correlated with IFC work quality ratings, particularly for project preparation work quality.

Explaining Project Performance

Several factors besides COVID-19 negatively affected IFC's investment project performance. IFC investment projects in the CY20–22 cohort suffered from unfavorable economic issues (23 percent of projects), high business risks (17 percent), and higher-than-expected competition (14 percent). Economic factors reduced demand for IFC client products and services and lowered the project companies' operational and financial performance compared with the projections at the Board approval stage. Financial sector projects dealing with high business risks moved away from lending to riskier segments to preserve capital. In the real sector, adverse business factors related to cyclicality, a downturn in the markets, or untested and flawed business models affected investment project performance. Higher-than-expected competition led to investment projects missing operational targets and contributed to reduced operating margins and profitability.

Despite these challenges and those posed by COVID-19, IFC's private sector clients showed remarkable resilience, adaptability, and flexibility. For example, in the financial sector, most of IFC's clients contracted their loan portfolio and focused on asset quality issues. Many real sector project companies implemented cost-saving initiatives to increase efficiency and shore up margins. Others invested quickly in information technology solutions to facilitate remote work. Many companies rolled out online versions of their business lines, particularly companies in the higher education and food and consumer retail sectors. IFC health care, food packaging, and consumer retail sector investment projects adapted their

services and products to respond to COVID-19–related demand changes. As such, "capable sponsors" positively affected project performance. Consequently, IFC investment projects were also successful when IFC complemented its financing with advisory projects to increase the capacity of sponsors and clients. The provision of advisory projects is one of the examples of how IFC can deliver nonfinancial additionality.

IFC has no formal procedures for modifying investment projects' development objectives to adapt to changing market conditions after Board approval. By their nature, private sector projects must be financially sustainable to survive in a competitive environment. At the same time, IFC projects are meant to achieve development objectives and comply with IFC's environmental and social performance standards. If needed, IFC can restructure the terms of investment financing agreements with clients and reschedule loan repayment schedules and clients can adapt their products and services to changing market conditions, such as those caused by the COVID-19 pandemic. However, the original development objectives, indicators, and targets cannot be modified to reflect the changes in market conditions, as neither IFC processes nor the Anticipated Impact Measurement and Monitoring (AIMM) framework consider formal changes of development objectives or targets after approval by the Board.

IFC investment project objectives were highly outcome oriented; however, outcome achievement rates were low, and measurement shortcomings persisted. Overall, all reviewed IFC investment projects established project-level outcomes, and a majority (74 percent) expected to achieve market-level outcomes, in line with the IFC 3.0 strategy. IFC fully achieved 45 percent of investment project outcomes, including both project-level and market-level outcomes, and partially achieved 22 percent. IFC investment projects that achieved more of their intended outcomes achieved higher development outcome ratings. Most of these investment projects in the *RAP 2023* cohort were not subject to an AIMM assessment at their approval and continued to be monitored in the Development Outcome Tracking System. In many cases, IFC or IEG used other available information sources and validated the outcomes, but 8 percent of intended outcomes could not be verified because of a lack of evidence.

Multilateral Investment Guarantee Agency

Project Performance

MIGA guarantee projects' development outcome ratings remained stable over the last six years but were slightly lower over the last three years, partially due to pandemic-related market challenges. MIGA guarantee projects in the *RAP* cohort were exposed to the pandemic for 27 percent of their active project lives. As a result, FY20–22 MIGA guarantee projects operated in a relatively more challenging operating environment during the pandemic. On a six-year rolling basis over FY17–22, MIGA's overall development outcome success ratings remained stable, with 72 percent of guarantee projects rated satisfactory or better. However, ratings were lower on a three-year rolling basis over FY20–22, reflecting the more challenging operating environment. That said, there were some delays in the delivery of some MIGA self-evaluations, which limited the number of guarantee projects analyzed in the *RAP* cohort. Therefore, this *RAP*'s analysis provides only limited and preliminary insights on the pandemic's effects on MIGA guarantee projects. Like IFC, MIGA has no formal procedures for restructuring development-related objectives or outcome targets during project implementation or crises.

The performance gap between guarantee projects in IDA and blend countries and those in non-IDA countries largely stayed the same in FY17–22. The performance of IDA and blend projects continued to be lower than that of non-IDA projects, with 64 percent rated satisfactory or better for development outcome in FY17–22. In contrast, guarantee projects in non-IDA countries maintained satisfactory or better ratings for 76 percent of projects in FY17–22. That said, MIGA's overall development outcome ratings in guarantee projects in FCS countries were on par with those of projects in non-FCS countries, with 70 percent of projects rated satisfactory or better in FY17–22.

MIGA work quality was rated lower than satisfactory in half of guarantee projects in FY17–22 and continued to exhibit shortcomings. Sixty percent of MIGA guarantee projects for the six-year rolling average over FY12–17 were rated satisfactory or better, but the share fell to 51 percent in FY16–21 and to 50 percent in FY17–22. The decline was even more evident when looking at three-year rolling averages, which fell from 56 percent rated satisfactory or

better in FY15–17 to 48 percent in FY19–21 and just 43 percent in FY20–22. MIGA work quality rating was correlated with the development outcome rating in 75 percent of guarantee projects in FY17–22.

MIGA achieved high success rates in carrying out its expected role and contribution. The share of guarantee projects with satisfactory ratings for MIGA's role and contribution was 88 percent in FY17–22—the same level as in FY16–21. MIGA's role and contribution ratings were generally high across the entire portfolio, including those in FCS, and IDA and blend countries. MIGA accomplished its expected role and contribution in almost 90 percent of guarantee projects. MIGA's role and contribution was most significant in environmental and social areas and risk reduction.

Explaining Project Performance

Pandemic- and non-pandemic-related factors undermined MIGA project implementation and performance. COVID-19–related lockdowns and economic downturns affected guarantee projects in the public transportation and energy sector by reducing consumer demand for these services. Other factors besides COVID-19 also negatively affected MIGA's project performance. Cost overruns and construction delays as well as foreign exchange issues were the most common adverse factors. However, some MIGA guarantee projects were able to adapt to the challenging economic landscape. For example, some MIGA hospital projects played an active role in assisting governments in meeting the new medical demands posed by COVID-19. More specifically, capable sponsors and a favorable legal and regulatory framework helped MIGA guarantee projects effectively adapt to implementation challenges.

Much like IFC, MIGA guarantee project objectives were highly outcome oriented, despite low outcome achievement rates and a lack of appropriate results measurement indicators and evidence. All MIGA guarantee projects pursued project-level outcomes, and 81 percent pursued at least one foreign investment–level outcome. However, the reviewed projects fully achieved 50 percent and partially achieved 22 percent of the outcomes defined at approval. This affected project ratings as the *RAP 2023* confirmed that achieving more intended project development outcomes led to higher development outcome ratings. Moreover, 69 percent of guarantee project development

outcomes were not tracked by MIGA because of a lack of indicators, which, after other supplemental verification, prevented the validation of 10 percent of expected development outcomes.

Future Directions for the World Bank Group

World Bank

Strengthen project capacity to identify and mitigate risks during project preparation, especially the risk of low implementing agency capacity.

Risk management by the World Bank's project teams and the technical capacity of implementing agencies were key factors in successful project implementation. Indeed, the weak capacity of implementing agencies emerged as the predominant underlying risk in projects that failed to adequately identify and mitigate risks. This underscores the need for World Bank project teams to conduct comprehensive risk assessments and develop robust mitigation strategies that prioritize capacity risks, especially in countries where local capacity limitations are common. This future direction aligns with the *RAP 2022* proposal to strengthen country programs' ability to assess implementation capacity risks.

Continue improving M&E as both an adaptation and accountability tool.

The World Bank took a proactive approach to adapt and restructure projects as needed during the COVID-19 crisis by closely monitoring projects' progress and identifying emerging challenges. M&E frameworks also provided sufficient evidence on project achievements. Thus, there is a valuable opportunity to scale up project monitoring, adaptation, and restructuring into postpandemic contexts and, more generally, beyond crisis scenarios. This will help maximize the resilience and performance of World Bank projects. Nevertheless, there are still areas in which the World Bank can continue to improve the M&E frameworks for greater accountability. In particular, the World Bank could enhance these frameworks' ability to measure institutional capacity outcomes in line with the outcome orientation agenda. This future direction is consistent with *RAP 2021*, which shows that not all projects with institutional strengthening objectives have adequate indicators to measure them.

International Finance Corporation

Improve the delivery of IFC additionality in difficult markets to enhance investment project outcomes. Difficult markets include those in FCS, Africa, and IDA and blend countries, in particular. We found that IFC additionality success ratings were particularly low in a large share of investment projects in these markets. The IFC 3.0 strategy aims to ramp up its investment program in these challenging markets. Higher realized IFC additionality in these challenging markets will make it more likely for IFC investment projects to achieve their objectives. IFC can add value to projects in these markets in several ways. For example, IFC delivers tailored financing but can also increase its provision of industry expertise and capacity-building advisory services, improve corporate governance, and enhance the environmental and social standards and practices of clients. Improving the delivery of IFC additionality would require IFC to adopt a proactive approach to ensure that additionality promises made at approval, particularly nonfinancial additionalities, are fulfilled and properly monitored during the investment project's life.

Further strengthen the selection of indicators and the measurement and tracking of intended development outcomes of investment projects. These measures would facilitate the monitoring of project development outcome progress and better reflect actual achievement. *RAP 2021* highlighted the challenges in measuring development outcomes, particularly at the market level, and this *RAP* showed that these challenges continue to be an issue. We found that monitoring data were not available for many intended development outcomes of IFC investment projects in the *RAP 2023* cohort. As such, IFC has an opportunity to improve its design and implementation of monitoring indicators to ensure that they can measure and track the achievement of intended project outcomes of investment projects. This would require IFC to provide clear definitions and sources for chosen indicators and ensure that clients have the capacity to measure them. That said, the investment projects in the *RAP 2023* cohort predate the rollout of IFC's AIMM framework, which requires IFC to track all project claims until the AIMM target year, which could improve some of these monitoring issues. IFC confirmed that it has increased the use of standardized indicators, improved regular monitoring, and engaged in an ongoing effort to establish a new data

platform for data tracking and reporting for investment projects approved under the AIMM system. Appropriate implementation of these measures could result in improvements in measurement and tracking of intended outcomes, although IEG has not yet been able to systematically validate these claims as very few IFC investment projects approved under the AIMM framework have been evaluated so far.

Multilateral Investment Guarantee Agency

Enhance project preparation work quality to strengthen the performance of MIGA guarantee projects. We found that MIGA work quality was rated lower than satisfactory in half of guarantee projects. MIGA could undertake more comprehensive project risk assessments, estimate detailed operational and financial projections with clear targets, and account for stricter downside scenarios. These up-front actions would help project teams enhance the awareness or understanding of potential project risks, consider mitigation mechanisms, and set clear project expectations. In public-private partnership projects, MIGA could identify foreseeable macroeconomic developments, such as local currency depreciations that can increase the government's financial obligations, and assess the risks from these developments, for example, whether the government is willing or able to pay the increased obligations to reduce their sustainability risks. According to MIGA, project risk assessments have recently improved, and the current Impact Measurement and Project Assessment Comparison Tool (IMPACT) framework incentivizes project teams to mitigate risks to the extent possible. However, IEG has not been able to validate these claims because MIGA's guarantee projects approved under the IMPACT framework have not yet been subject to evaluation.

Strengthen measurement and tracking of intended development outcomes, particularly at the foreign investment level. These measures would facilitate the monitoring of project development outcome progress, would better reflect actual achievement, and would be especially helpful for tracking the achievement of intended foreign investment–level outcomes. MIGA could accomplish this by better defining its project development objectives, selecting relevant indicators to measure outcomes, and establishing appropriate mechanisms to gather results evidence and development

impact data. This suggestion is in line with the findings of *RAP 2021*, which noted that many MIGA guarantee projects lacked sufficient evidence to rate project outcomes; however, MIGA's evidence collection has improved in recent years.

Management Comments

Management of the World Bank welcomes the Independent Evaluation Group (IEG) report *Results and Performance of the World Bank Group 2023* and thanks the team for incorporating comments provided. This is the 13th annual *Results and Performance of the World Bank Group (RAP)* report, and management welcomes the overall positive findings of performance at the project level and the various steps taken to address operational disruptions during the COVID-19 pandemic. The report's findings provide valuable insights for both project preparation and adaptive management during implementation. Management also welcomes the institution-specific suggestions for the future.

World Bank Management Comments

Overall Comments

Management is pleased with the overall findings of the report, including the improvements across all three elements of the *RAP* analysis—ratings of outcomes, Bank performance, and monitoring and evaluation (M&E). The average project outcome ratings are now at their highest since fiscal year (FY) 2012. This is noteworthy as the overall size of World Bank commitments is bigger than at any time in the past—the World Bank delivered record financial commitments in FY23: 66 percent over the precrisis average during FY14–19. This suggests that the incremental delivery has happened in parallel with staff attention to quality of projects under implementation. Bank performance ratings for development policy financing (DPF) projects, entailing both design and implementation, have improved since FY21 and are at the highest since FY12. At the same time, investment project financing (IPF) projects and Programs-for-Results projects have maintained their ratings between FY21 and FY22, including for quality at entry and quality of supervision. Efficacy ratings of intended development outcomes have shown statistically significant improvements during recent years. M&E quality ratings have shown consistent improvements, including notably in countries affected by fragility, conflict, and violence (FCV). While year-on-year changes

are important, it is of note that the overall trend in recent years across most indicators used in the *RAP* has been positive and steady.

Management welcomes the report's recognition of the effectiveness of proactive measures undertaken in response to the COVID-19 pandemic. These measures included project restructuring and adaptive implementation, which helped improve project performance—the report notes that "restructurings increased from an average of 1.9 per project in the prepandemic cohort to 2.6 in the *RAP 2023* cohort" (32). The report's validation of the benefits of timely project restructuring is particularly useful as teams consider measures to continually focus on outcomes: "projects that made course corrections earlier in the project cycle had a higher likelihood of achieving their intended development outcomes" (xii). Management also acknowledges the role played by clients to maintain project performance and results. Steps such as expedited decision-making and streamlined government procedures contributed to the quality of implementation by implementing agencies and project implementation units and underpinned timely project restructuring.

Management continues to be attentive to addressing project implementation challenges during this period of ongoing multiple crises. While being appreciative of the report's conclusion "that overall project performance did not suffer is a testament to the resilience and adaptability of project teams" (92), management is watchful of any potential downturn in performance as projects with extended exposure times to the pandemic and other compounding shocks are included in subsequent *RAP* reports.

Other Comments

Management is pleased with the report's acknowledgment of proactive measures taken to improve M&E ratings, which are now at their highest since FY12. Refining M&E methodologies, revising indicators, adjusting targets through restructuring, and gathering supplemental evidence on project achievements helped temper protractive challenges related to the inadequacy of indicators, unrealistic targets, lack of data collection methodology, and attribution issues. Together, these helped projects ascend from moderate to substantial efficacy ratings and achieve intended outcomes—enabling course corrections bases and providing improved evidence on project achievements.

Management will continue strengthening M&E data quality and systems, including through periodic deep-dive reviews of Implementation Status and Results Reports and Implementation Completion and Results Reports.

Management notes the disaggregated analysis of the extent of outcome orientation in results frameworks and the progress to date. Based on an analysis of 4,808 indicators in the *RAP 2023* cohort of projects, the report found that 40 percent of the project development objective indicators measured outcomes, 46 percent measured intermediate outcomes, and only 12 percent measured outputs. The report points to the scope for further improving outcome orientation for the intermediate results indicators, which management will learn from. The analysis did not find a significant association between a project's indicator level (that is, outcome or output) and its efficacy ratings and explains that this might be because "other lower-level indicators demonstrate that projects completed intended activities and that these activities would plausibly contribute to the achievement of intended development outcomes, as outlined by the project's theory of change" (37). Management notes this as evidence that tools such as the theory of change are playing a role in strengthening the lines of sight between project contributions and intended development outcomes. As part of the work on the new Corporate Scorecard, management is committed to strengthening the results architecture of the World Bank. The outcome-oriented focus of the scorecard will be reflected in the approach to be used for developing project-level indicators. Management is also examining the appropriate way of aligning incentives, capacity, and institutional systems with the outcome-orientation approach of the new scorecard.

Future Directions

Management concurs with the report's suggestion to continually strengthen client capacity to identify and mitigate risks during project preparation, especially the risk of low implementing agency capacity. This suggestion is particularly salient with the continued expansion of support to clients facing greater uncertainties. The Systematic Operations Risk-Rating Tool calibration exercise has enabled teams to identify and mitigate residual risks and helped management focus on high and substantial risk projects. As part of the evolution's workstream related to operational effectiveness

and efficiency, management is looking at ways to strengthen country capacity and institutions and systems. It is also applying a risk-based approach to project preparation and implementation to focus attention that is proportional to risks and where it is most needed, such as low-capacity environments. The nature of implementation agency capacity risks varies significantly across Regions and sectors and between FCV and non-FCV contexts. Where the risk is acute, management is committed to addressing underlying factors through close monitoring, capacity building, commissioning third-party expertise to supplement systems, and use of technology. Experience also shows that relying only on ex ante risk assessment may not be sufficient, as counterpart capacity to implement World Bank–financed operations tends to be weaker in early phases of the project cycle. Updating of risk assessment would therefore need to be part of early course correction and adaptive management.

Management concurs with the report's suggestion to continue improving M&E as an adaptation and accountability tool. This agenda is advancing further with work on the new Corporate Scorecard, which will help manage and course correct with evidence and report results at scale. As part of the scorecard's implementation plan, management is planning to improve data quality, impact measurement, and investments in World Bank skills and M&E client capacity on data quality and data management. The scorecard introduces results narratives as a core component under each outcome area. Applying rigorous methodologies such as process tracing and contribution analysis, the results narratives will help articulate World Bank's contributions to the enabling institutional and policy environment. The expanded use of impact evaluations in the planned Global Challenge Programs is another measure that will strengthen impact measurement. All these measures taken together will improve the results data quality and support midcourse corrections based on evidence, in line with the outcome-orientation agenda.

International Finance Corporation Management Comments

International Finance Corporation (IFC) management welcomes the flagship IEG report *RAP 2023*. Deep dives on (i) the evolution and relationship of project development outcomes to project performance ratings and (ii) the

influence of the COVID-19 pandemic on project performance are particularly helpful. IFC would appreciate a deeper analysis in future to better understand whether the documented correlation among IFC work quality, additionality, and development outcomes extends to a causal relationship and how development outcomes are affected.

Management notes that COVID-19–related effects have started to materialize and to affect performance. The pandemic contributed to a challenging operating environment for both investment and advisory projects, resulting in depressed and changing patterns of demand, reduced access to capital, rising bankruptcies, and persistent uncertainty. Multiple exogenous shocks further exacerbated pandemic-induced economic downturns. IFC welcomes IEG's assessment of factors that influence the implementation and performance of projects within this context, especially the report's observations that the performance of investment projects in African and fragile and conflict-affected situations (FCS) countries was challenged mostly by (i) adverse economic factors, (ii) high business risks, and (iii) limited technical expertise and track record of sponsors and clients. As delivery ramps up in these challenging contexts, management is committed to strengthening IFC's project preparation and M&E capabilities, while recognizing that significant factors remain outside its control.

Management acknowledges the continued weak outcome ratings for IFC investments in Africa, in countries classified as FCS, in countries eligible for International Development Association (IDA), and in World Bank countries and notes that adverse macroeconomic factors and high business risks are key drivers of performance for this group. After last year's *RAP 2022* Board of Executive Directors discussion, IFC undertook an internal review, and many of the drivers identified by IEG in its deep dive resonate with the review's findings. The IFC deep dive also showed that projects in Africa and IDA-FCS contexts were particularly affected by the challenging economic and operating environment.[1] In IFC's analysis, external factors and risks that underlay weak development outcome ratings included (i) project- or industry-specific factors (fall in prices, weak demand for services or products, sector-specific regulatory and licensing challenges); (ii) unfavorable external macrofactors (economic slowdown, conflict, political turmoil, local currency depreciation, and infectious disease outbreaks); and (iii) sponsor or management issues

(lack of management attention, turnover, inexperience, and weak relationship with a sponsor). Manufacturing and infrastructure projects were mostly affected by project or industry factors, while financial institutions were vulnerable to macroeconomic and sponsor issues. FCS projects were exposed to multiple factors that increased the severity of their impact.

While acknowledging that more work needs to be done to support development outcomes, management would also like to highlight IFC's sustained progress in enhancing delivery in African, IDA, and FCS countries to meet strategic objectives in these priority markets. In FY23, IFC invested $10.4 billion across 41 countries in Africa, the highest ever annual commitment in the continent. Long-term finance reached $6.9 billion ($3.8 billion of own account and $3.1 billion core mobilization), and short-term finance and trade and supply chain in the region amounted to $3.5 billion. For the 17th Replenishment of IDA (IDA17) and FCS, the project count reached 41 percent of IFC's total (surpassing the IFC Corporate Scorecard target of 39 percent), illustrating an increase in delivery of projects with high expected development outcomes. Short-term finance commitments reached $7.5 billion in IDA17 and FCS countries (68 percent of total short-term finance) and $3 billion in IDA17 low-income countries and FCS countries (27 percent of total short-term finance) in FY23. Moreover, IFC management has taken deliberate actions over the past years to bolster successful outcomes in African, IDA, and FCS countries. These include the merger of upstream and advisory teams in the regions to better align efforts to create the conditions necessary for private sector investment through client and project preparation work; adding experienced, senior resources in the field; and increasing environmental and social capacity in country offices. IFC has also established dedicated programs such as the Africa Fragility Initiative, focused on developing and implementing investment, upstream, and advisory programs in 32 fragile countries in Africa, and a Joint IFC-UNHCR Initiative to enable private sector solutions in the forced displacement context.

Management notes the report's comments around IFC additionality and highlights two points. First, the report's comparison of anticipated and realized additionality for select projects in challenging markets relies on nomenclature that is IEG's interpretation of the categories defined in IFC's 2018 Revised Additionality Framework. The framework was not applied to

investment projects in the *RAP 2023* cohort, which predates the framework's rollout. Second, management acknowledges that IFC is comparatively less successful in realizing nonfinancial additionality than financial additionality and agrees that IFC needs to be more deliberate about nonfinancial additionality. To this end, IFC is focused on providing industry expertise, capacity-building advisory, and better monitoring of delivery of additionality.

Management appreciates IEG's analytics and the finding on outcome types noting a high level of outcome orientation in project objectives but disagrees with the assertion of deficient tracking of project-level outcomes. Though the previous M&E system—Development Outcome Tracking System—was retired in FY20 and replaced by the Anticipated Impact Measurement and Monitoring (AIMM) system since FY18, results tracking for all active investment projects in the *RAP* cohort takes place and is ongoing (both systems were active until FY20). While most pre-AIMM vintage projects are not AIMM assessed during portfolio monitoring, development impact indicators of all investment projects were monitored and assessed by one of these systems. Deficiencies of the Development Outcome Tracking System in capturing claims have been addressed in the AIMM system. Further, IFC is stepping up analytical work in assessing ex post outcomes for specific projects and programs. This will complement the ongoing ex ante AIMM analysis and yield deeper insights into development outcomes and drivers of project success. To this end, IFC is hiring new staff with expertise in conducting impact evaluations.

Management notes that, in line with the report's Future Directions section, IFC has already mainstreamed AIMM into its investment operations, strengthening the measurement and tracking of intended development outcomes of investment projects. *RAP 2023* covers projects that predate AIMM. The subsequent adoption and full implementation of the AIMM system has already helped ensure increased use of standardized indicators and their regular monitoring, with an ongoing effort to establish a new data platform for data tracking and reporting, in line with the renewed vision and mission for the World Bank Group. The AIMM framework enables IFC to not only connect financing with quantifiable development outcomes but to also communicate the impact goals to the Board, stakeholders, and clients. As part of the ongoing enhancements to the AIMM framework, IFC is launching the AIMM

Navigator—a new tool designed to create a more seamless impact rating and data management process, centralize IFC's development impact and additionality data, and bring more efficiency to the impact data collection and reporting process.

The report observes that IFC investment projects have no formal procedures for modifying their original development objectives, indicators, and targets to adapt to changing market conditions. Management appreciates IEG noting this weakness in IFC's approach and, in collaboration with IEG, will explore ways to address it.

Multilateral Investment Guarantee Agency Management Comments

The Multilateral Investment Guarantee Agency (MIGA) welcomes IEG's *RAP 2023* report and finds it useful and important. MIGA thanks IEG for the productive engagement during the report's preparation. In particular, MIGA finds the *RAP 2023* valuable as the first systematic reporting of World Bank project performance during the COVID-19 pandemic.[2] We also note that this *RAP* updates and enhances the outcome orientation analysis of World Bank projects covered in the *RAP* two years ago. The *RAP 2023* also applied innovative use of machine learning for the International Bank for Reconstruction and Development, the IDA, and IFC assessments. MIGA is hopeful that the possible synergies and efficiency gains of this approach can be applied to the analysis of MIGA projects going forward, and MIGA stands ready to work with IEG to operationalize this enhancement, if this would be helpful. MIGA also observes that the scope of the *RAP 2023* did not include an assessment of the effectiveness of collaboration across the World Bank—a significant aspect of the work pertaining to the "Evolution of the World Bank" and the "new playbook." While the subject was touched on in the previous *RAP*, reflecting in part a focus on more country-level evaluation evidence, MIGA would find it useful if IEG were to have more systematic and regular coverage of the One World Bank approach in future *RAP*s.

MIGA welcomes *RAP 2023*'s confirmation that MIGA maintained its historically high development outcome success rate of 72 percent. Also, MIGA welcomes IEG's observation that the development outcome success ratings

of projects in FCS and non-FCS countries remain the same. This finding is important given projects in FCS countries are riskier and more challenging. Conversely, *RAP 2023* reported projects in IDA and Blend countries are less successful than in non-IDA countries. Pandemic-related challenges and related higher country risk environments are headwinds for MIGA, as evidenced by IEG's observation that over a quarter of the MIGA projects of the cohort evaluated in RAP2023 experienced COVID-19 pandemic effects—the highest percentage among the three institutions of the World Bank. MIGA is increasingly active in supporting projects in more challenging environments with higher associated risks, including fragile settings. Indeed, MIGA's overall portfolio in IDA and Blend countries has increased from 24 percent in FY19 to 41 percent in FY23. IEG also acknowledged MIGA's positive role and contribution (that is, referring to outcomes associated with company behavior changes due to MIGA's participation), with satisfactory ratings at 88 percent in FY17–22, the same level as in FY16–21.

The *RAP 2023*'s assessment of MIGA's work quality indicates that evaluated projects had a lower proportion of satisfactory and above ratings in the recent cohort compared with the earlier periods. In the Future Directions section, it suggests that MIGA "enhance project preparation work quality to strengthen the performance of MIGA guarantee projects" (98). MIGA considers IEG's observations in the Future Directions section related to its assessment of front-end work quality to be based mainly on the assessment of MIGA's front-end work quality from 5 to 10 years ago. Since then, MIGA has made significant changes, partly due to previous IEG observations in this area of MIGA's work quality and lessons, and MIGA has learned from previous discussions of project evaluations with IEG.

In this regard, it is important to recognize that the FY17–22 projects evaluated for *RAP 2023* entered MIGA's portfolio in about FY14–18, corresponding to a period of strong growth in guarantee issuance and product innovation, including new areas of risk-taking and testing new approaches and instruments. During this period, MIGA was in the process of adapting its work product to reflect the new challenges and lessons learned, including in the latter portion of the period by working to specify expected development impact based on the new instrument innovations and contexts. These experiences and learnings were essential for subsequent improvements,

which culminated in the launch of the Impact Measurement and Project Assessment Comparison Tool (IMPACT) framework, MIGA's ex ante development outcome assessment tool, piloted in FY19 and fully launched in FY20.

Another indication of less-than-satisfactory rated MIGA work quality is associated with IEG's observation that scenario analysis would highlight the riskiness of the project business and that MIGA could undertake more comprehensive project risk assessments, estimate detailed operational and financial projections with clear targets and account for stricter downside scenarios. However, this observation is based on the historical cohort of evaluated projects assessed for *RAP 2023* and does not represent MIGA's current approach to its project work. MIGA's current project work is heavily oriented toward assessing project risks, detailing downside scenarios, and mitigating risks where feasible. Most project documents now contain downside scenarios when appropriate; this development in MIGA's project work is partly a result of prior IEG observations on this point from which MIGA has learned and implemented relevant changes. Moreover, MIGA's IMPACT framework provides a likelihood assessment of development outcomes, which brings attention to development outcome risks and incentivizes teams to mitigate these risks to the extent feasible and, in the process, potentially achieve higher IMPACT scores. However, identifying risks is not the same as successfully mitigating risks, which needs to be balanced against the costs and practicality of mitigating risks, which has a bearing on a project's bankability and feasibility. MIGA emphasizes calculated and appropriate risk taking to support projects in the most difficult contexts.

On the discussion of outcome orientation, MIGA is also pleased by IEG's assessment that MIGA guarantee projects were focused on higher-level outcomes as envisaged by their specific intended outcomes, which was a main objective in introducing the IMPACT framework itself. There is a specific observation by IEG that MIGA should strengthen the measurement and tracking of intended development outcomes, particularly at the foreign investment–effects level. MIGA agrees with this observation, and MIGA is at the early stages of making changes to how MIGA tracks projects in relation to the IMPACT framework, especially (as also noted by IEG) for the achievement of intended foreign investment–level outcomes. These reforms will also be helpful in the context of launching the new World Bank

Corporate Scorecard, for which MIGA is actively engaged in discussions with International Bank for Reconstruction and Development and IDA and IFC colleagues. MIGA hopes to show progress in this area in forthcoming Project Evaluation Results and future *RAP* reports.

[1] The review consisted of a deep-dive analysis of 84 investment projects (out of 298 validated by the Independent Evaluation Group) in African, IDA, and fragile and conflict-affected situation countries with unsuccessful development outcome ratings during calendar years 2015–21. The four constituent dimensions of development outcome—project business success, economic sustainability, private sector development, environmental and social effects—were analyzed. Twenty-four projects in this sample were in and fragile and conflict-affected situation countries.

[2] "World Bank" in this Management Response refers the International Bank for Reconstruction and Development, the International Development Association, the International Finance Corporation, and the Multilateral Investment Guarantee Agency.

1 | Introduction

This *Results and Performance of the World Bank Group* examines the evolution of the World Bank Group's project-level results and performance, analyzing the development outcomes and factors that have shaped project implementation and performance, especially in the context of the COVID-19 pandemic.

The methodological approach of this report is underpinned by three principles: continuity, innovation, and symmetry. The report embraces continuity by building on previous work, innovation by incorporating innovative data and evaluation methods, and symmetry by applying them in a balanced and consistent manner across different Bank Group institutions, while taking into account each institution's different evaluation and rating methods.

This is the 13th annual *Results and Performance of the World Bank Group* (*RAP*) report. The *RAP* comprehensively reviews the development effectiveness of the World Bank Group through evidence gathered by the Independent Evaluation Group (IEG). Specifically, this report assesses the results and performance of the World Bank, which includes 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).

RAP 2023 focuses on three main evaluation questions: (i) How did IEG ratings change over time at the project and country levels across the various Bank Group institutions? (ii) What has been the evolution of development outcomes pursued, measured, and achieved at the project level, and what is the relationship of outcomes to project performance ratings? (iii) What factors affected the Bank Group projects' implementation and performance in the COVID-19 pandemic context? These evaluation questions aimed to respond to specific areas of inquiry that were highlighted by Bank Group management and the Board of Executive Directors during *RAP* consultations.

This *RAP* updated project performance rating trends from the past 10 years—fiscal years (FY)12–22—to answer the first evaluation question. These updated performance ratings covered World Bank lending projects, IFC investment and advisory projects, and MIGA guarantee projects (see chapters 2, 3, and 4, respectively).[1] Although this is the first *RAP* with a significant number of projects that took place during the pandemic, many projects were already at an advanced stage of implementation when the pandemic began, minimizing the severity of their exposure. Therefore, this *RAP* does not fully capture the impact of COVID-19 on projects' performance, and it is likely that project cohorts in the coming years will reflect the full consequences of the COVID-19 pandemic.

Results and performance across Bank Group institutions are not strictly comparable because of different project evaluation and rating methodologies. For instance, the World Bank projects employ an objective-based methodology to derive project performance ratings. These ratings summarize the World Bank's self-evaluation and IEG's validation narratives into categories or values that enable aggregation across operations.[2]

Similarly, IFC's advisory project performance ratings are derived from an objective-based methodology, which establishes minimum thresholds for rating and assessing these projects' effectiveness. By contrast, evaluation systems and performance ratings for IFC's investment projects and MIGA's guarantee projects largely rely on a benchmark-based methodology. This benchmark-based methodology aligns with good practice standards for evaluating private sector projects, as established by the Evaluation Cooperation Group of multilateral development banks (ECG 2011). See box 1.1 for a description of the main performance ratings for each institution and appendix A for more details on each institution's ratings and evaluation methodology.

Box 1.1. Main Performance Ratings in World Bank Investment Projects, IFC Investment and Advisory Projects, and MIGA Guarantee Projects

World Bank

outcome: The extent to which a project efficiently achieved, or was expected to achieve, its relevant objectives. The outcome rating brings together three underlying dimensions: relevance, efficacy (objectives achievement), and efficiency. It is rated on a six-point scale: highly satisfactory, satisfactory, moderately satisfactory, moderately unsatisfactory, unsatisfactory, and highly unsatisfactory.

relevance: The extent to which a project's objectives are consistent with current World Bank Group country strategies at the time of project closing. It is rated on a four-point scale: high, substantial, modest, and negligible.

efficacy: The extent to which a project achieves, or was expected to achieve, its objectives, taking into account the objective's relative importance. The project's achievement of each individual objective is assessed based on the concept of "plausible causality." Efficacy ratings also reflect an assessment of the results framework's validity and use complementary data and evidence on the achievement of intended results. Both the efficacy of each individual objective and overall efficacy in achieving the project development objective are rated on a four-point scale: high, substantial, modest, and negligible.

(continued)

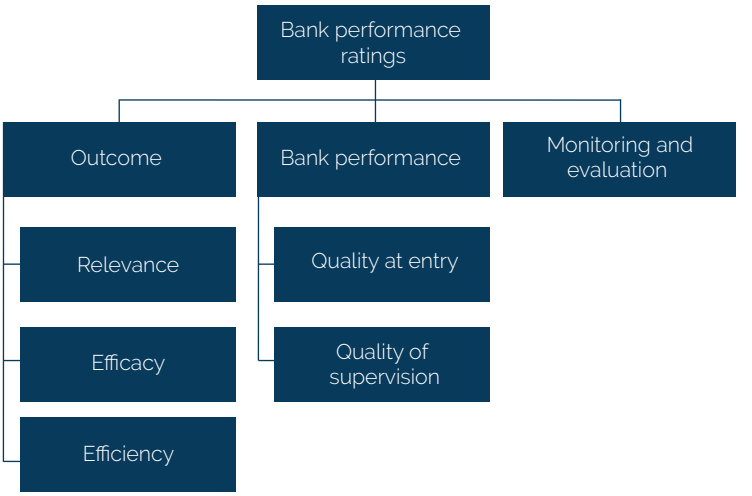
Box 1.1. Main Performance Ratings in World Bank Investment Projects, IFC Investment and Advisory Projects, and MIGA Guarantee Projects (cont.)

efficiency: How economic resources and inputs are converted to results. Efficiency ratings indicate whether the costs involved in achieving project objectives were reasonable compared with project benefits and recognized norms (value for money). It is rated on a four-point scale: high, substantial, modest, and negligible.

Bank performance: The extent to which World Bank services ensured quality project design and supported effective implementation through appropriate supervision in the achievement of development outcomes. Bank performance and its two constituent elements—quality at entry and quality of supervision—are rated on a six-point scale: highly satisfactory, satisfactory, moderately satisfactory, moderately unsatisfactory, unsatisfactory, and highly unsatisfactory.

monitoring and evaluation (M&E) quality: The quality of the design and implementation of project M&E and the extent to which M&E results are used to improve performance. M&E quality is assessed at the project level and includes M&E design, implementation, and use. It is rated on a four-point scale: high, substantial, modest, and negligible.

Figure B1.1.1. Performance Ratings in World Bank Investment Projects



Source: Independent Evaluation Group.

Note: This is the ratings structure for investment project financing and Program-for-Results; development policy financing has a slightly modified ratings structure (see appendix A).

(cont.)

Box 1.1. Main Performance Ratings in World Bank Investment Projects, IFC Investment and Advisory Projects, and MIGA Guarantee Projects (cont.)

IFC

Investment Projects

development outcome: Synthesizes a project's performance across four subdimensions considering its impact on each affected stakeholder group: project business performance, economic sustainability, environmental and social effects, and private sector development. It is rated on a six-point scale: highly successful, successful, mostly successful, mostly unsuccessful, unsuccessful, and highly unsuccessful.

IFC additionality: The benefit or value addition IFC brings to a project that a client would not otherwise have. It is rated on a four-point scale: excellent, satisfactory, partly unsatisfactory, and unsatisfactory.

IFC investment outcome: The extent to which IFC has realized at the time of evaluation and expects to realize over the remaining life of the investment, the project's loan income, equity returns, or both based on what was expected at approval. It is rated on a four-point scale: excellent, satisfactory, partly unsatisfactory, and unsatisfactory.

IFC work quality: IFC's operational performance, including in relation to environmental and social aspects, with respect to precommitment work in screening, appraising, and structuring, and supervision and administration work after the Board of Executive Directors' approval of the project and the subsequent commitment. It is rated on a four-point scale: excellent, satisfactory, partly unsatisfactory, and unsatisfactory.

Advisory Projects

development effectiveness: Synthesizes a project's performance across five indicators: strategic relevance, output achievement, outcome achievement, impact achievement, and efficiency. It is rated on a six-point scale: highly successful, successful, mostly successful, mostly unsuccessful, unsuccessful, and highly unsuccessful.

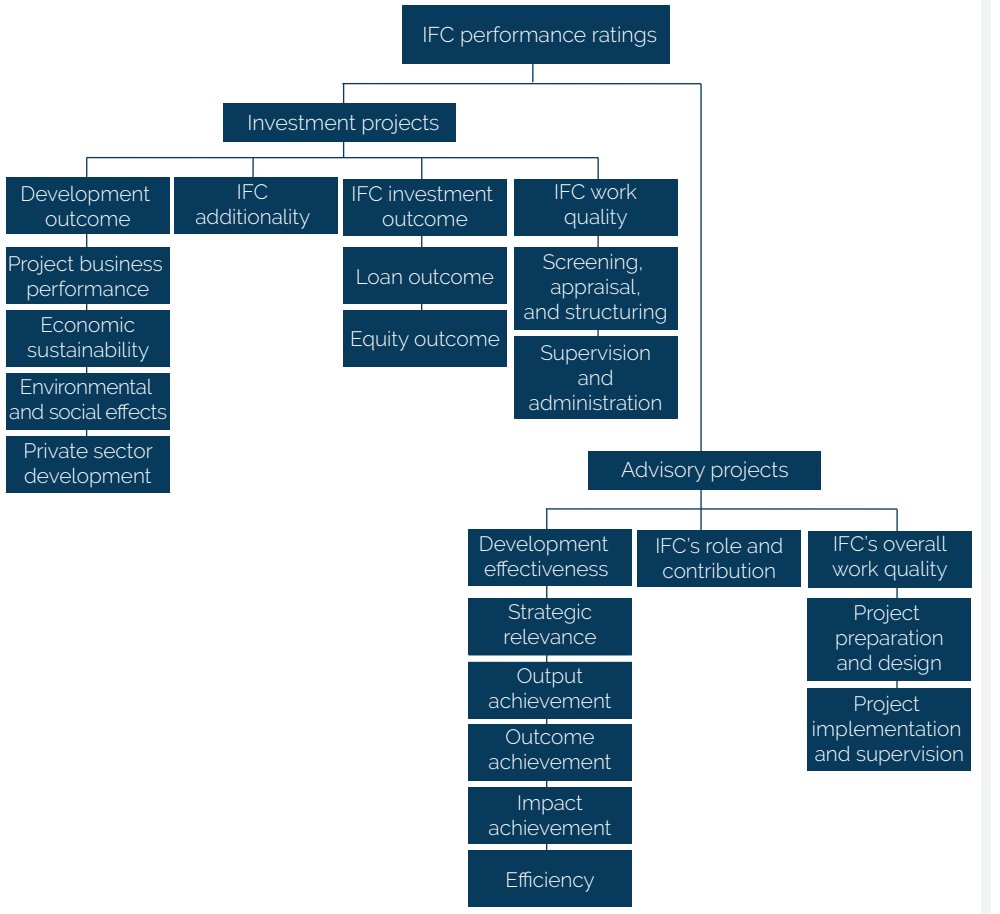
IFC's role and contribution: The extent to which IFC added value or made a special contribution to the advisory project. It is rated on a four-point scale: excellent, satisfactory, partly unsatisfactory, and unsatisfactory.

(continued)

Box 1.1. Main Performance Ratings in World Bank Investment Projects, IFC Investment and Advisory Projects, and MIGA Guarantee Projects (cont.)

IFC's overall work quality: The extent to which IFC services ensured quality at entry and supported effective implementation, through appropriate supervision and execution, toward the achievement of development objectives. IFC work quality and its two dimensions—project preparation and design, and project implementation and supervision—are rated on a four-point scale: excellent, satisfactory, partly unsatisfactory, and unsatisfactory.

Figure B1.1.2. Performance Ratings in IFC Investment and Advisory Projects



Source: Independent Evaluation Group.

Note: IFC = International Finance Corporation.

(continued)

Box 1.1. Main Performance Ratings in World Bank Investment Projects, IFC Investment and Advisory Projects, and MIGA Guarantee Projects (cont.)

MIGA

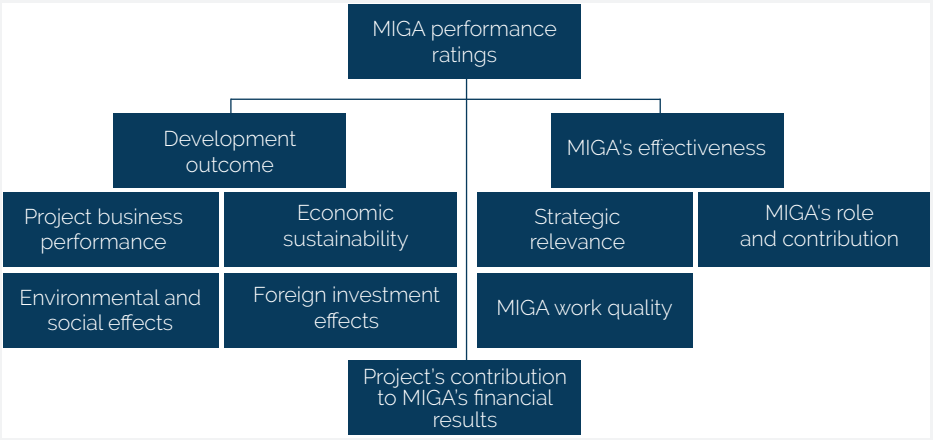
development outcome: Measures performance across four indicators considering its impact on each affected stakeholder group: project business performance, economic sustainability, environmental and social effects, and foreign investment effects. It is rated on a six-point scale: highly successful, successful, mostly successful, mostly unsuccessful, unsuccessful, and highly unsuccessful. Until fiscal year 2019, the ratings were based on a four-point scale: excellent, satisfactory, partly unsatisfactory, and unsatisfactory.

MIGA's effectiveness: Synthesizes MIGA's performance across three indicators: project strategic relevance, MIGA's role and contribution, and MIGA work quality. It is rated on a four-point scale: excellent, satisfactory, partly unsatisfactory, and unsatisfactory.

MIGA's role and contribution: The benefits and value added that MIGA brings to the client, project, or political risk insurance industry. It is rated on a four-point scale: excellent, satisfactory, partly unsatisfactory, and unsatisfactory.

MIGA work quality: Assesses due diligence and underwriting processes, including of risk assessment and mitigation and monitoring after the issuance of the guarantee. It is rated on a four-point scale: excellent, satisfactory, partly unsatisfactory, and unsatisfactory.

Figure B1.1.3. Performance Ratings in MIGA Guarantee Projects



Source: Independent Evaluation Group.

Note: MIGA = Multilateral Investment Guarantee Agency.

We carried out in-depth analyses to answer the second and third evaluation questions. For the World Bank, the in-depth analyses covered all investment project financing (IPF) operations that closed between March 2020 and April 2022 (273 projects) and that had IEG ratings as of December 2022. These projects were approved between 2003 and 2021. The analysis compared this FY20–22 cohort (the *RAP 2023* cohort) with a prepandemic cohort of 398 IPF projects that closed and completed their Implementation Completion and Results Reports (ICRs) between July 2017 and February 2020. For IFC, the in-depth analysis focused on 170 investment operations that were included in IFC's Expanded Project Supervision Report (XPSR) program for calendar years (CY)20–22 and completed by December 2022. The *RAP* team then compared this *RAP* 2023 cohort with 265 prepandemic projects evaluated and validated as part of the XPSR program for CY17–19. For MIGA, the in-depth analysis covered 16 MIGA guarantee projects evaluated and validated by December 2022 from the Project Evaluation Report program for FY20–22.

Our methodological approach was guided by three principles: continuity, innovation, and symmetry. These principles collectively shaped our analysis and allowed for standardized comparisons of the Bank Group's results and performance across time. At the same time, the approach incorporated innovative research elements in a balanced and consistent manner for each Bank Group institution. The differences in project evaluation and rating methodologies for each institution were accounted for in the symmetrical application of these approaches.

This *RAP* builds on the research from previous *RAP*s, thereby ensuring continuity and creating symmetry in *RAP* analyses. We extended the *RAP 2021* analysis of development outcome types, allowing longer-term comparative analysis across the FY12–14, FY17–20, and FY20–22 periods.[3] Moreover, we deepened the inquiry started in *RAP 2021* by linking the achievement of specific project outcomes to individual project development outcome ratings. This added value to previous analyses because IFC's and MIGA's project development outcomes are rated at the project and subdimension levels but not at the specific outcome level. This is also the first *RAP* in which all projects took place, at least partly, during the pandemic. As such, this *RAP* expands on the limited findings of previous *RAP*s on the impact of COVID-19 on the Bank Group's results and performance.

This *RAP* also introduced several innovative analyses that apply similar methods in a balanced and consistent manner across Bank Group institutions. First, we conducted content analyses of World Bank projects' self-reported implementation factors and of IFC investment and MIGA guarantee projects' key performance factors, especially within the context of COVID-19. For World Bank projects, an adapted version of the DeCODE (Delivery Challenges in Operations for Development Effectiveness) taxonomy was used to classify the implementation factors, whereas for IFC investment and MIGA guarantee projects, the *RAP* team leveraged the Project Insights taxonomy of performance factors developed by IEG. Second, we used supervised machine learning exercises to analyze the factors affecting World Bank and IFC investment projects' implementation and performance over previous years. Third, we explored the reasons for low efficacy ratings in World Bank projects by using novel data from the Implementation Completion and Results Report Review (ICRR) system. Specifically, this analysis aimed to determine whether lower efficacy ratings are indicative of a failure to achieve well-defined indicators or a failure in providing sufficient evidence. Fourth, we investigated for the first time the impact of the COVID-19 pandemic on World Bank project restructuring patterns.

The report is organized into five chapters. Chapter 1 is this introduction. Chapter 2 focuses on the results and performance of the World Bank, chapter 3 on the results and performance of IFC, and chapter 4 on the results and performance of MIGA. Each of these three chapters looks at the exposure of projects to COVID-19, describes the presence of a sample selection bias, examines project performance rating trends, analyzes factors affecting project implementation and performance, explores the evolution of intended development outcomes, and assesses the validity of results framework indicators in measuring these outcomes. Chapter 2 also explores World Bank projects' restructuring patterns. Chapter 5 provides concluding remarks for each institution and future directions to potentially take. These chapters are complemented by a set of appendixes that provide additional data and supporting information.

[1] The ratings analysis in the *Results and Performance of the World Bank Group 2023* report is based on the independent ratings of the Independent Evaluation Group, unless otherwise specified.

[2] The evaluation methodology for development policy financing projects has changed. The new methodology was developed starting in mid-2020 and finalized in June 2022. See table A.1 for a comparison of the old and new methodologies of the Implementation Completion and Results Report Review.

[3] The project outcome-type data that were created by *Results and Performance of the World Bank Group 2021* and used for comparative analysis only include projects closed in fiscal years (FY)12–14 and FY17–20 (second quarter).

2 World Bank Results and Performance

This is the first *Results and Performance of the World Bank Group* with a substantial number of closed projects affected by the COVID-19 pandemic during implementation; however, these projects still had a limited exposure to the pandemic, and successful projects are likely overrepresented in the sample.

The World Bank's overall project performance was not undermined by the effects of the COVID-19 pandemic and the Russian invasion of Ukraine. World Bank projects encountered pandemic-related and other obstacles that hindered implementation, despite those projects being at an advanced stage of implementation before the onset of the pandemic.

The World Bank's adaptive management and project restructuring during the pandemic contributed to improved project performance.

Improvements in the World Bank's monitoring and evaluation quality facilitated project adaptation and contributed to providing sufficient evidence of projects' achievements.

This chapter presents the World Bank performance rating trends of projects that were closed in FY12–22 and evaluated by June 30, 2023. This chapter also analyzes the factors that affected the implementation and performance of IPF, including project restructuring patterns. In addition, it explores the evolution of intended project-level development outcomes and assesses the validity of results framework indicators for measuring these outcomes. It also examines the associations between the validity of indicators and efficacy ratings across projects' intended development outcomes.

Project Exposure to the COVID-19 Pandemic and Sample Selection Bias

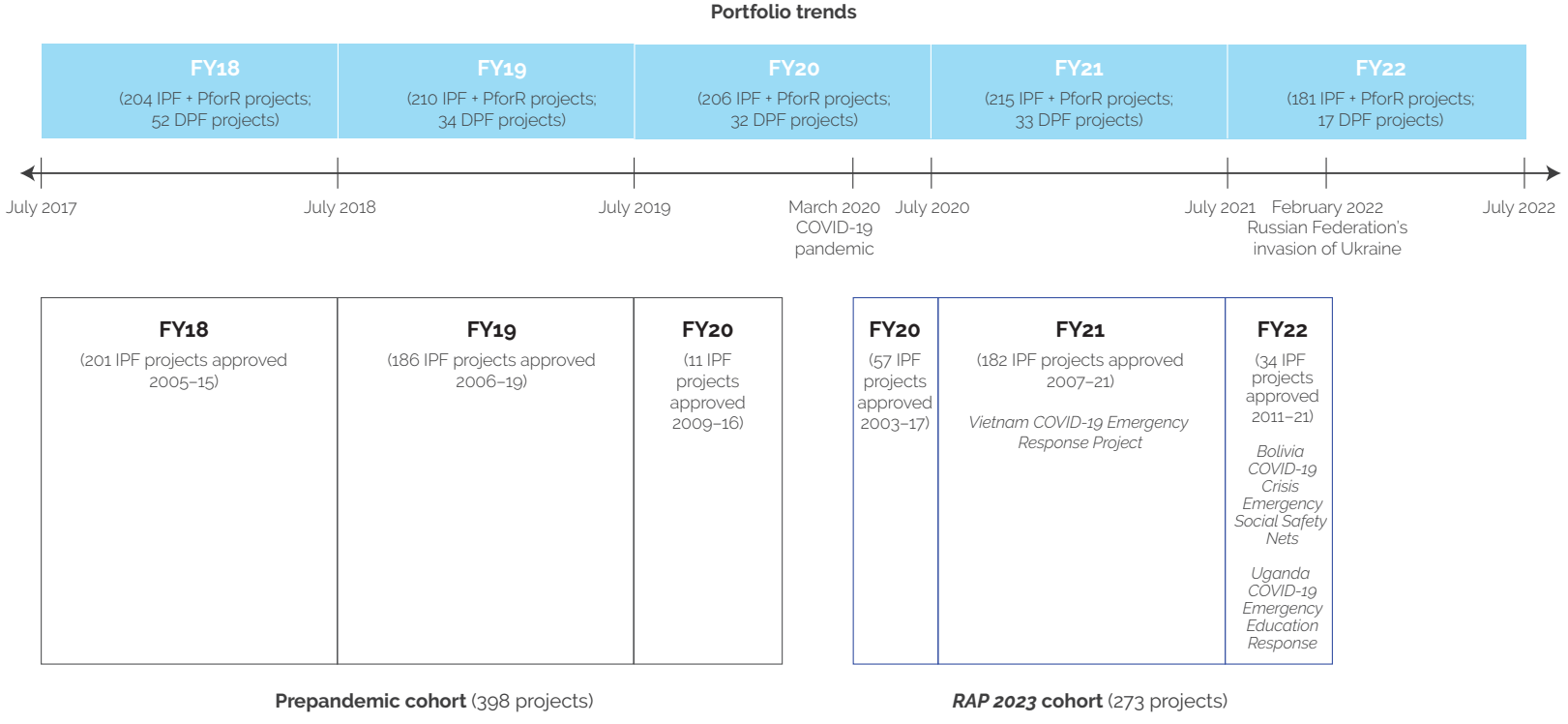
This is the first *RAP* with a substantial number of projects under implementation during the COVID-19 pandemic. Previous *RAP*s had limited findings on COVID-19 because very few projects in those *RAP*s took place during the pandemic. For example, *RAP 2020* found that the pandemic did not disrupt the World Bank's self-evaluation or IEG's validation processes (World Bank 2020). *RAP 2022* found that among the 10 closed projects specifically designed in response to the pandemic, 7 received satisfactory outcome ratings and 3 received moderately satisfactory outcome ratings (World Bank 2022a). The *RAP 2023*'s analysis of rating trends contained projects that operated during the pandemic, including 684 lending operations that were closed in FY20–22 and evaluated by IEG by June 30, 2023. This *RAP*'s in-depth analyses focused on 273 IPF projects (*RAP 2023* cohort) that were closed between March 2020 and April 2022 and evaluated by IEG by December 2022—all of which operated during the pandemic. The in-depth analyses also included a prepandemic cohort of 398 projects closed in FY18–20 before the pandemic began for comparison purposes. Figure 2.1 illustrates the difference in composition among the overall World Bank portfolio, the FY20–22 *RAP 2023* cohort, and the FY18–20 prepandemic cohort.

Projects in the *RAP 2023* cohort still had limited exposure to the COVID-19 pandemic. Many projects were already at an advanced stage of implementation when the pandemic began, minimizing the severity of their exposure. On average, only 14 percent of the total life span of the cohort's projects was during COVID-19.[1] Approximately half of the cohort's projects were exposed for less than 12 percent of their project life span,

and 90 percent were exposed for less than 23 percent of their life span. In fact, some projects reported that the pandemic had a limited impact on the quality, nature, or extent of implementation because these projects were already nearing completion when the pandemic began (see box A.1).

In addition, the *RAP 2023* cohort is susceptible to a sample selection bias. This overrepresentation of successful projects with higher ratings arises because projects that complete ICRs and ICRRs shortly after closing tend to have higher ratings than projects with delayed ICRs and ICRRs. In other words, the longer it takes to complete the ICR and ICRR, the lower the project ratings. This pattern also applies to the rating trends and occurred in previous years (figure A.3, panel a). However, the *RAP 2023* cohort is even more likely to have an overrepresentation of projects with higher ratings because of the evaluation cutoff date of December 2022. The early cutoff date was set to accommodate the time required for the *RAP*'s in-depth analysis and its data collection. An exploration of the latest Implementation Status and Results Report ratings on progress toward achieving project development objectives (PDOs) in FY22 shows that projects with completed ICRRs—which were therefore included in this *RAP*'s analysis of rating trends—have higher average Implementation Status and Results Report ratings than projects with completed ICRs but in-progress ICRRs and even higher ratings than projects with uncompleted ICRs (figure A.3, panel b). Therefore, rating trends should be interpreted carefully as they are likely to change downward in the future. This is especially true as more projects with extended exposure times to COVID-19 are incorporated into the project rating trends (see appendix A for more details on the limitations of the data).

Figure 2.1. Composition of the Overall Portfolio in Rating Trends, the Prepandemic Cohort, and the *RAP 2023* Cohort



Source: Independent Evaluation Group.

Note: FY = fiscal year; DPF = development policy financing; IPF = investment project financing; PforR = Program-for-Results; *RAP* = *Results and Performance of the World Bank Group*.

Project Performance Rating Trends

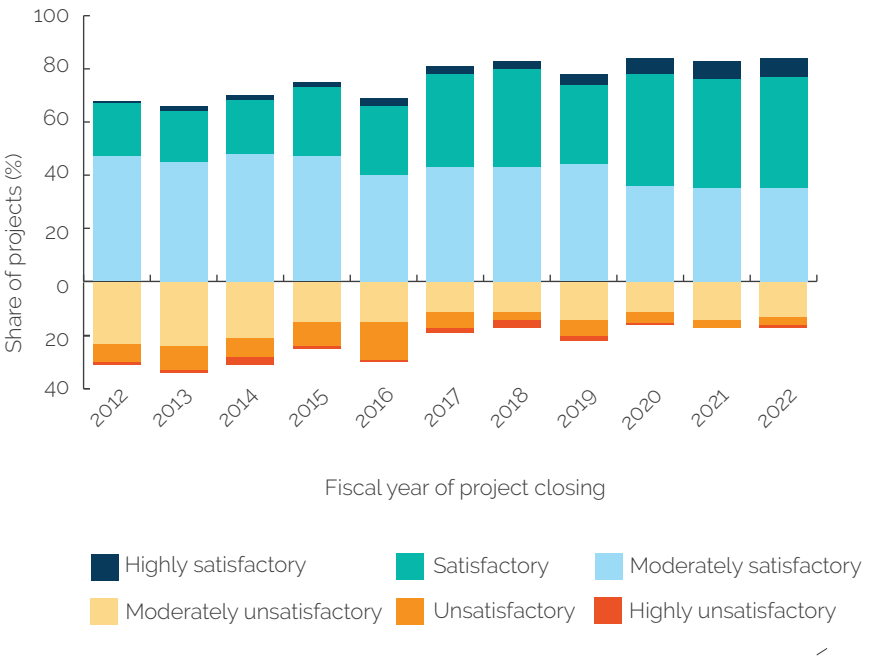
The World Bank's project outcome ratings remained high in FY22, despite the effects of the COVID-19 pandemic and the Russian invasion of Ukraine.[2] The average outcome rating of 181 IPF and Program-for-Results (PforR) projects in FY22 remained at 4.3 on a 6-point scale as in FY20 and FY21—the highest average since FY12—with the share of projects rated moderately satisfactory or above also staying constant at 83 percent between FY21 and FY22. Moreover, there was a slight improvement in the share of IPF and PforR projects rated satisfactory or above, increasing from 47 percent in FY21 to 49 percent in FY22 (figure 2.2, panel a). This pattern in outcome ratings indicates projects' resilience to the adverse global context across a large share of project subgroups—including by Region, Global Practice, country income level, and others—rather than being solely influenced by the portfolio's shift toward highly rated project subgroups (figure B.3). The average outcome rating for 17 development policy financing (DPF) projects in FY22 stayed at an average of 4.0 on a 6-point scale, with the share of projects rated satisfactory slightly increasing from 33 percent in FY21 to 35 percent in FY22. There was a fairly similar decline in the share of projects rated moderately satisfactory, decreasing from 45 percent in FY21 to 41 percent in FY22. Therefore, the share of DPF projects rated moderately satisfactory or above slightly declined from 79 percent to 76 percent (figure 2.2, panel b). This decline in DPF project ratings should be interpreted with caution because of the limited sample size—in FY21, there were only 33 DPF projects, and in FY22, the number decreased further to just 17 projects.

World Bank projects also maintained or improved their average Bank performance ratings. Bank performance ratings for IPF and PforR projects also stayed flat, with an average rating of 4.3 on a 6-point scale in both FY21 and FY22. Although the share of projects rated moderately satisfactory or above declined marginally from 87 percent in FY21 to 86 percent in FY22, the share of projects rated satisfactory or above actually increased from 39 percent in FY21 to 43 percent in FY22 (figure 2.3). The average quality-at-entry rating for IPF and PforR projects—a subcomponent of the Bank performance rating—also remained constant at 4.2 on a 6-point scale, with an increase from 42 percent of projects rated satisfactory and above in FY21 to 44 percent in FY22 but also a decrease from 82 percent of projects rated moderately satisfactory and above in FY21 to 75 percent in FY22. A decomposition analysis shows that,

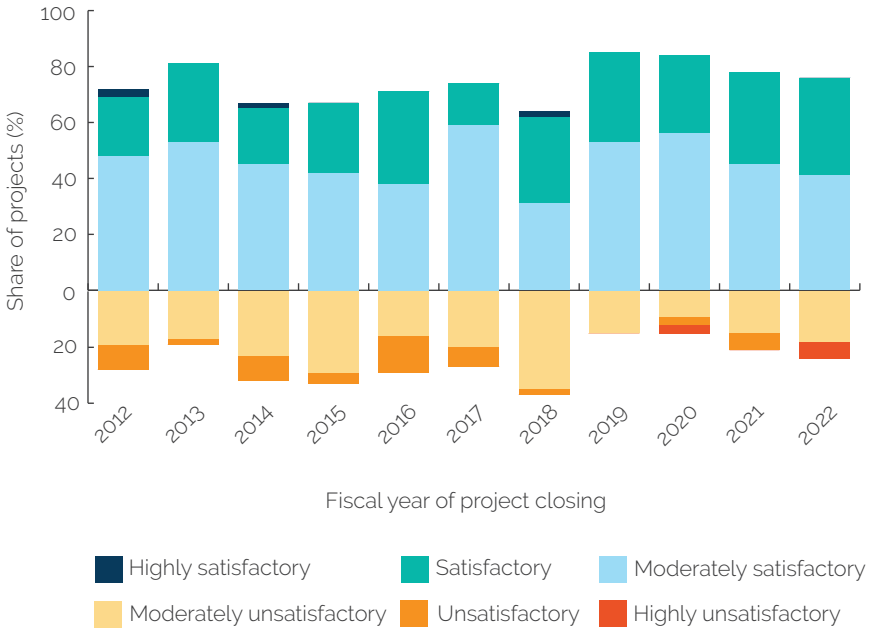
across World Bank Regions, the negative shift in quality-at-entry ratings is largely explained by a drop in project ratings in the South Asia Region, from 85 percent in FY21 to 60 percent in FY22. These quality-at-entry ratings in FY22 were not linked to project preparation challenges caused by COVID-19 because the vast majority of FY22 projects had been approved before March 2020 (figure B.6). Quality-of-supervision ratings—the other subcomponent of the Bank performance rating—also stayed constant at 4.6 on a 6-point scale, with the share of projects rated highly satisfactory increasing from 4 percent in FY21 to 8 percent in FY22 and the share of projects rated moderately satisfactory or above slightly decreasing from 92 percent in FY21 to 91 percent in FY22. Conversely, Bank performance ratings for DPF projects improved from an average rating of 4.3 in FY21 to 4.6 in FY22 on a 6-point scale, and the share of projects rated moderately satisfactory or above went up from 94 percent in FY21 to 100 percent in FY22 (figure 2.4). Design and implementation ratings, which replaced quality-at-entry and quality-of-supervision ratings in DPF projects, exhibited similar patterns, with design ratings increasing from 91 percent in FY21 to 100 percent in FY22 and implementation ratings increasing from 94 percent in FY21 to 100 percent in FY22.[3]

Figure 2.2. World Bank Project Outcome Ratings

a. Investment project financing and Program-for-Result projects



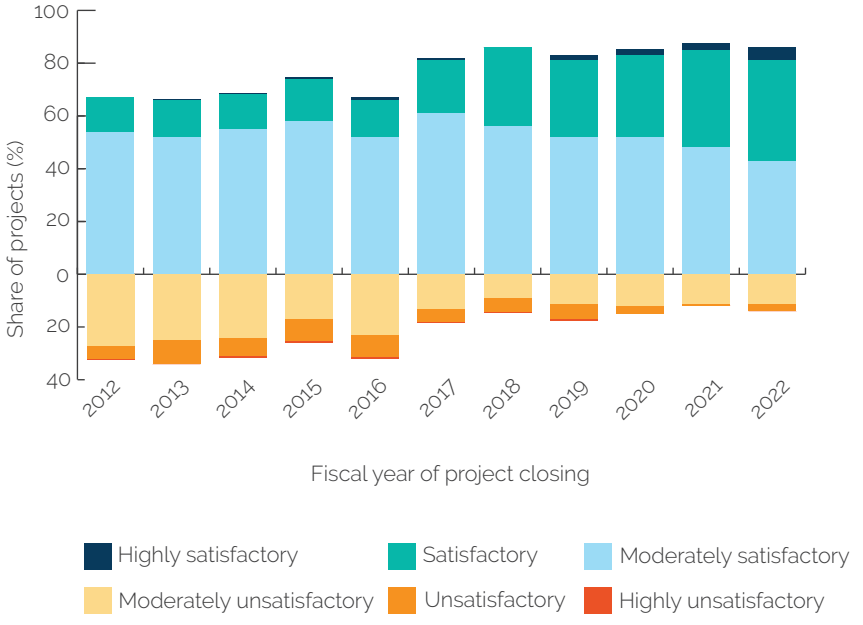
b. Development policy financing projects



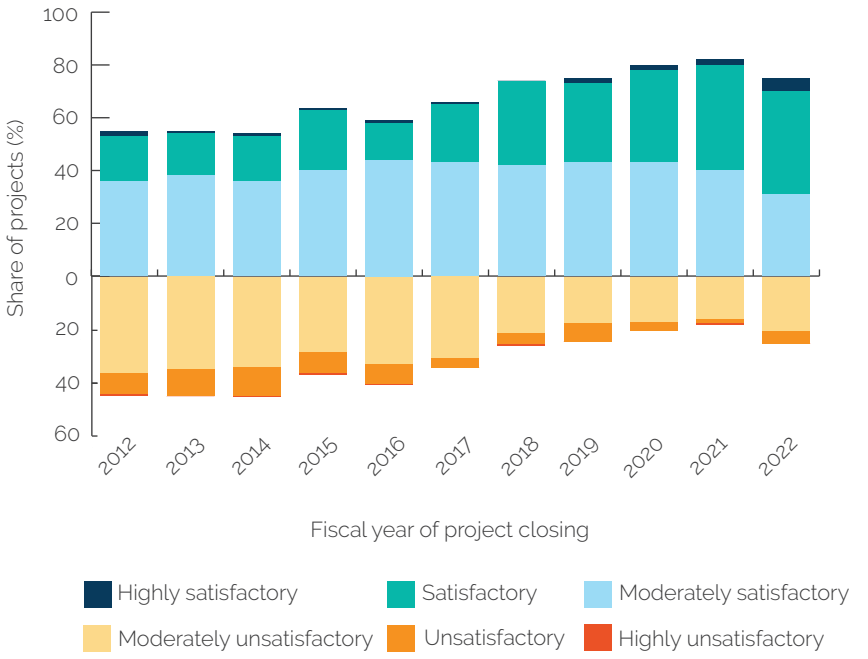
Source: Independent Evaluation Group.

Figure 2.3. Bank Performance, Quality at Entry, and Quality of Supervision for Investment Project Financing and Program-for-Results

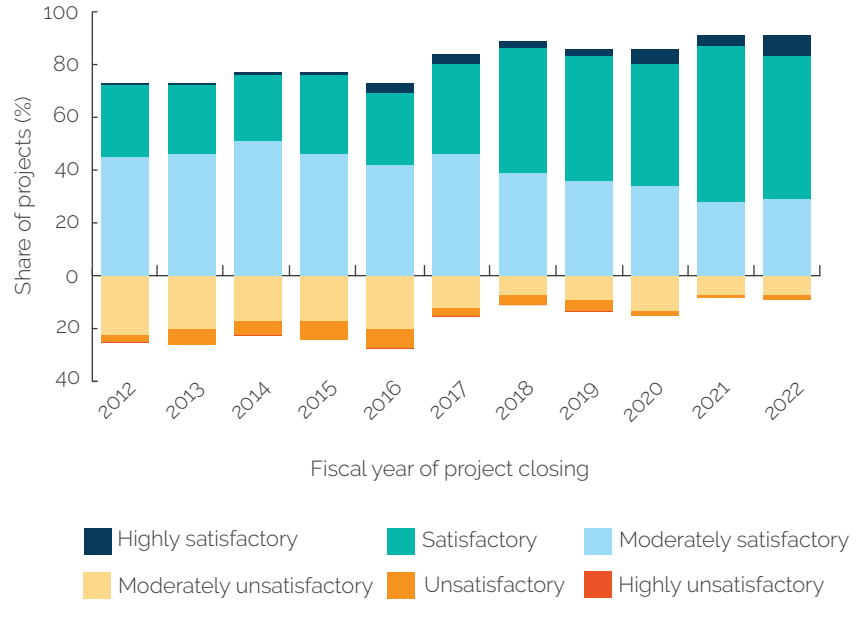
a. Bank performance



b. Quality at entry



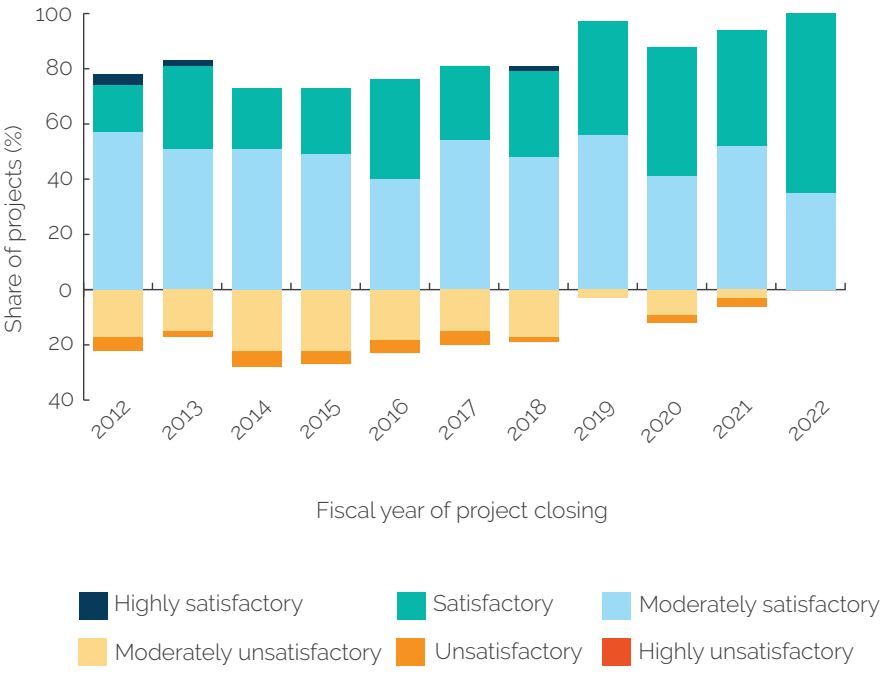
c. Quality of supervision



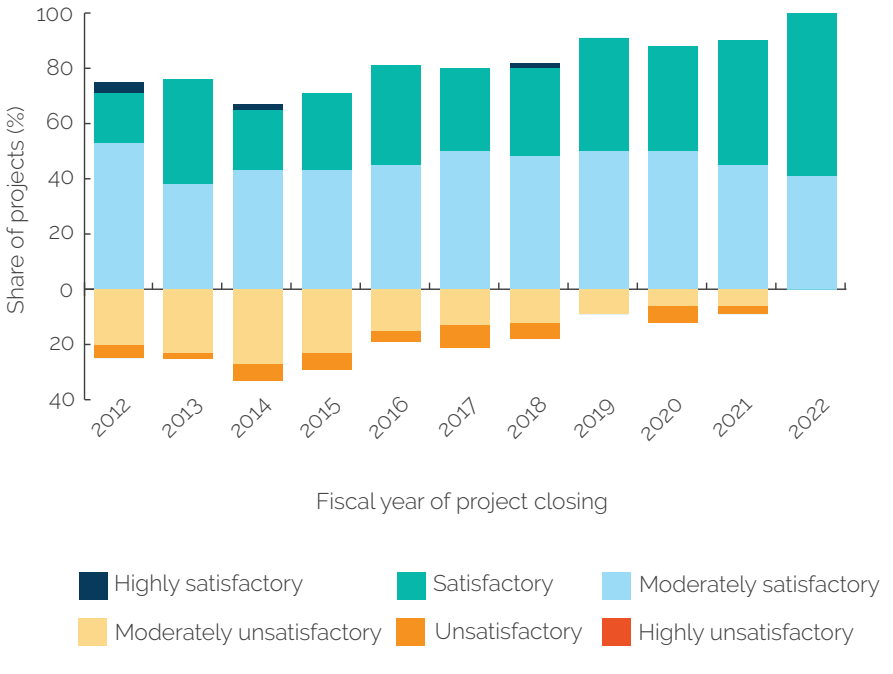
Source: Independent Evaluation Group.

Figure 2.4. Bank Performance, Design, and Implementation for Development Policy Financing

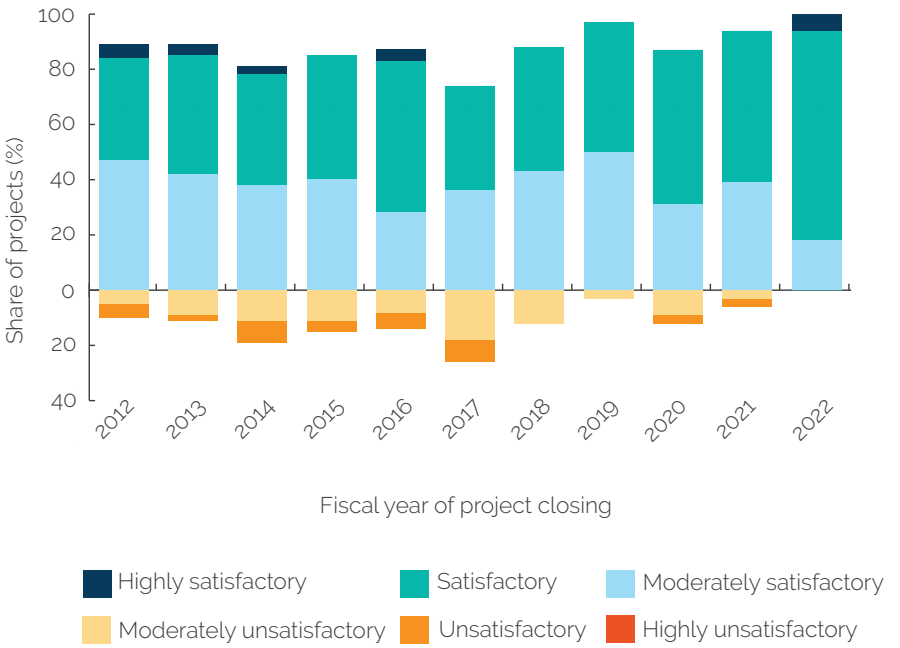
a. Bank performance



b. Design



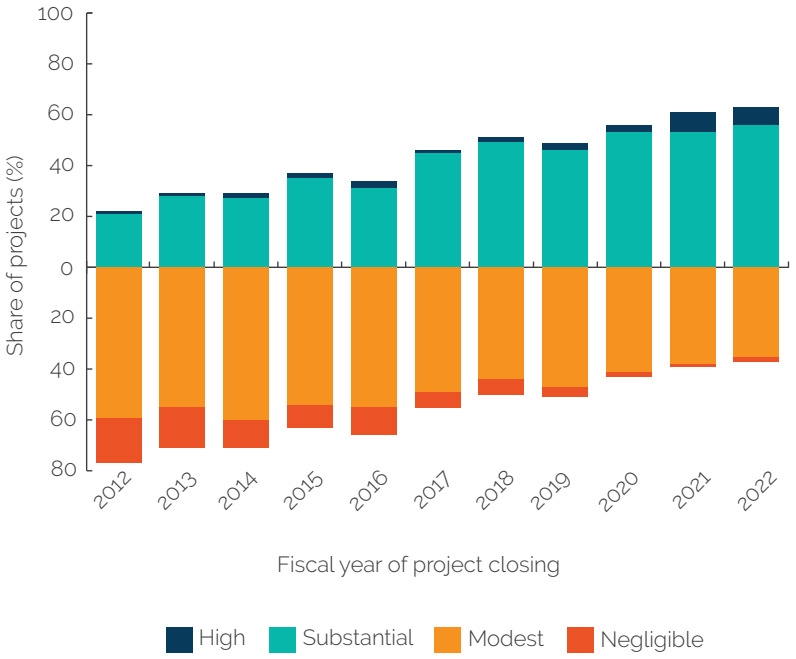
c. Implementation



Source: Independent Evaluation Group.

The World Bank's monitoring and evaluation (M&E) quality ratings have consistently improved. The share of IPF and PforR projects rated substantial or high in M&E quality increased from 60 percent in FY21 to 63 percent in FY22 (figure 2.5). This increase was driven by the improved ratings of the Infrastructure Practice Group, up from 37 percent to 56 percent, and the portfolio expansion of the high-performing Human Development Practice Group, which grew from 20 percent to 27 percent of the overall portfolio. M&E quality ratings in IDA fragile and conflict-affected situation (FCS) countries also significantly increased, from 48 percent to 60 percent. M&E ratings declined notably for Western and Central Africa, where the share of projects rated substantial or high dropped from 67 percent to 53 percent. The South Asia Region and the Europe and Central Asia Region had the most pronounced increase, with a growth of 15 percentage points in South Asia and of 11 percentage points in Europe and Central Asia (figure B.9).

Figure 2.5. Monitoring and Evaluation Quality Ratings for Investment Project Financing and Program-for-Results Projects



Source: Independent Evaluation Group.

Note: Monitoring and evaluation quality ratings for development policy financing are not reported because they have been dropped from Implementation Completion and Results Report Reviews under the new methodology.

Factors Affecting Project Implementation and Performance

The COVID-19 pandemic was the single most salient challenge facing projects during FY20–22. Despite the limited exposure time to the pandemic, 212 projects, or 78 percent, experienced implementation obstacles caused by the pandemic,[4] as reported by ICR documents (figure 2.6).[5] Lockdowns and mobility restrictions had adverse effects on countries' economic activity, leading to disruptions in services and public institution operations. Most projects reported implementation delays caused by supply chain shortages and other logistical challenges, which had an impact on civil works components of projects. The pandemic also led to the postponement of in-person project-related activities and, in some cases, the reallocation of project funds (box 2.1).

Box 2.1. The Impact of the COVID-19 Pandemic on Project Implementation

This *Results and Performance of the World Bank Group* uncovered the pandemic's specific underlying effects on projects' implementation. We conducted a content analysis of 443 extracts of Implementation Completion and Results Report text corresponding to the 212 projects identified with the epidemics factor (see figure 2.6). The underlying effects include the following:

Lockdowns, mobility restrictions, and economic downturn. The outbreak of the COVID-19 pandemic had significant repercussions in projects' implementation as countries declared states of emergency, imposed nationwide lockdowns, and implemented mobility restrictions, including border closures to curb the virus's spread. These measures had adverse effects on countries' economic activity, particularly for informal workers and poor households, and contributed to job losses and even permanent firm closures.

Disruption of services. The implementation of education projects was particularly affected, with widespread school closures disrupting ongoing and planned academic activities and leading to learning losses among students. The closure of technical and vocational education and training institutions in North Macedonia and Afghanistan affected practical training for students, making it challenging for small and medium companies to absorb them. Health and transport projects saw reduced service use, affecting the delivery of services supported by World Bank projects. Preventative health services witnessed a decline as people avoided health care facilities because of contagion fears. Transport projects experienced disruptions, with decreased travel and railway services leading to lower demand and interruptions. Colombia, for example, had an 85 percent drop in public transport demand.

Disruption of operations of institutions. In addition, across all Global Practices, World Bank projects reported that government agencies at the national and local levels faced temporary disruptions to their work schedules and operations, hampering interactions and active engagement with project stakeholders.

(continued)

Box 2.1. The Impact of the COVID-19 Pandemic on Project Implementation (cont.)

Slow-paced activities. The pandemic presented obstacles during the final stages of project implementation across all Global Practices, thus slowing down the pace of activities that in many cases led to the extension of project closing dates to compensate for the lost time during lockdowns.

Shortages in supply chain and logistics challenges delayed civil works. Supply chain and logistics challenges, resulting from lockdowns, border closures, and travel restrictions, caused delays in civil works, particularly in energy, transport, water, and urban projects.

Difficulty of in-person activities. Travel and mass gathering restrictions had a significant impact on project activities requiring physical interaction and mobility (such as training programs, workshops, and technical meetings), leading to the postponement, cancellation, or shift to virtual formats across all Global Practices. This also affected supervision and verification activities, including field missions, making it challenging for technicians and World Bank staff to monitor project progress. Nine projects reported that the pandemic hindered the collection of primary data and field visits, resulting in delayed project reports and the exclusion of certain result indicators from monitoring and Implementation Completion and Results Report preparation.

Reallocation of project funds. In addition, 15 projects reported that the COVID-19 crisis exerted pressure on government budgets and shifted priorities toward pandemic response efforts, leading to cancellation or redirection of project funds to mitigation measures.

Sources: Independent Evaluation Group.

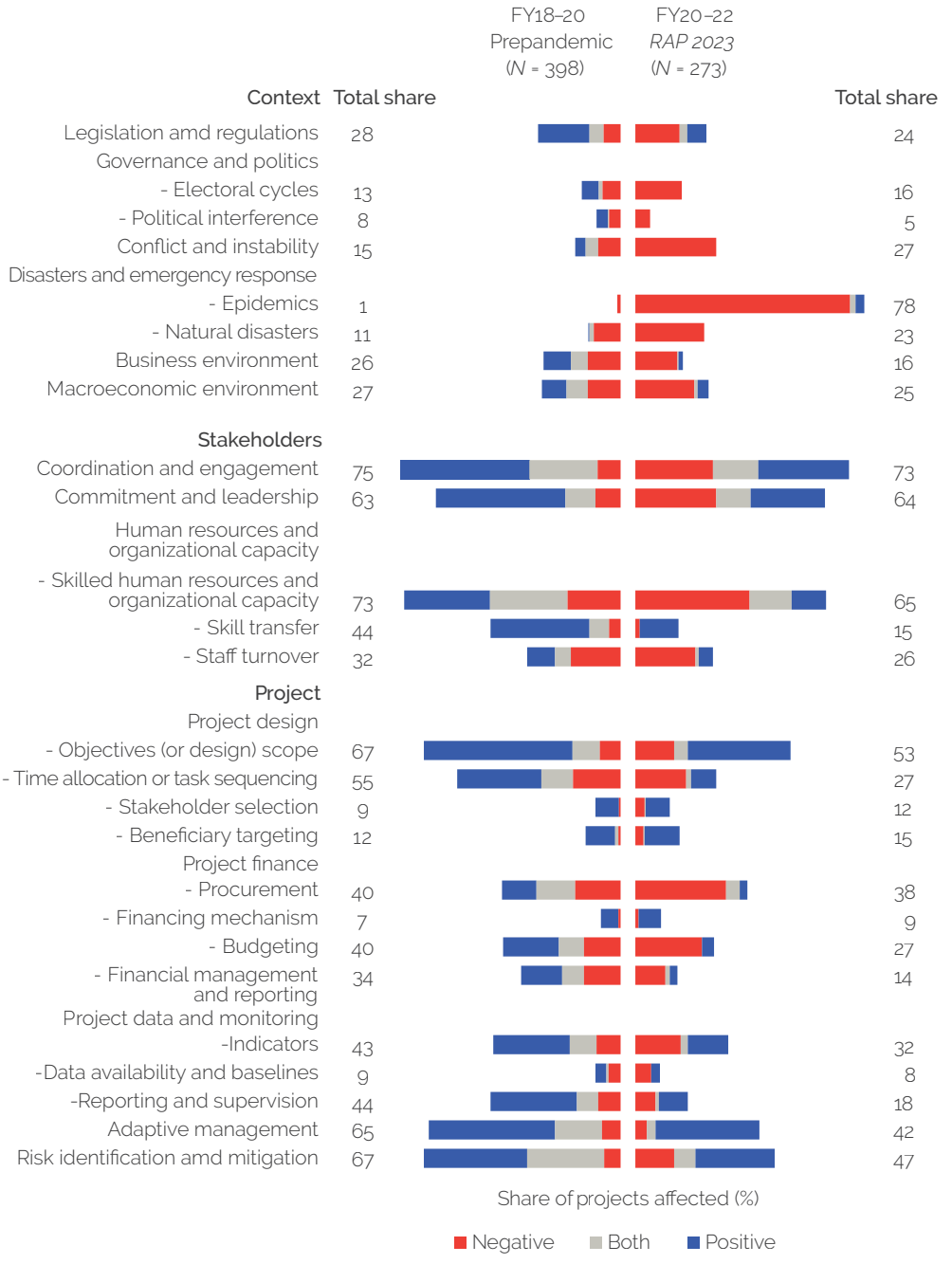
Countries' institutional capacity, procurement, and conflict and instability were other common challenges during project implementation. The low technical capacity of implementing agencies to execute and supervise work quality hindered the implementation of 39 percent of projects. Such weak institutional capacity was common in the South Asia and Eastern and Southern Africa Regions and in IDA and FCS countries.[6] About 31 percent of projects reported challenges with procurement management systems,

including delays and inefficient contract management. Procurement challenges were more prevalent in low-income countries, in the South Asia and Europe and Central Asia Regions, and in the Infrastructure and Sustainable Development Practice Groups. Conflict and instability, more prevalent in the Western and Central Africa Region and in IDA and FCS countries, also hindered the implementation of 27 percent of projects in the *RAP 2023* cohort.

By contrast, project scope, ex ante risk identification and mitigation, and adaptive management facilitated project implementation. Among project-related factors, 35 percent of projects highlighted that a realistic scope for objectives or strong overall project design had facilitated implementation. Project teams underscored adaptations to unforeseen circumstances as helping implementation in 35 percent of projects and ex ante risk identification and mitigation measures as helping in 27 percent. While across Regions projects tended to report on the adequacy of risk identification and mitigation measures for project implementation, projects in the Latin America and the Caribbean Region, in contrast, tended to report on the inadequacy or insufficiency of teams' risk identification and mitigation measures for successful project implementation (see figure D.6).

The inadequate identification and mitigation of institutional capacity risks emerged as a challenge in project implementation. Twenty-one of the 56 projects, or 38 percent, acknowledged the failure to adequately identify and mitigate risks and reported that weak implementing agency capacity was the most important implementation risk (figure 2.7; table 2.1). These projects commonly reported that the initial risk assessments conducted before project implementation were overly optimistic given the complexity of the project. Consequently, the proposed mitigation measures proved insufficient, leading to delays in project implementation. Moreover, 15 out of 21 projects also encountered implementation obstacles caused by the low technical capacity of implementing agencies, which is captured by the skilled human resources and organizational capacity subcategory. Consistent with this finding, *RAP 2022* also found that World Bank country programs were less adept at assessing institutional capacity risks (World Bank 2022a).

Figure 2.6. Factors Affecting Project Implementation: A Comparative Analysis



Source: Independent Evaluation Group.

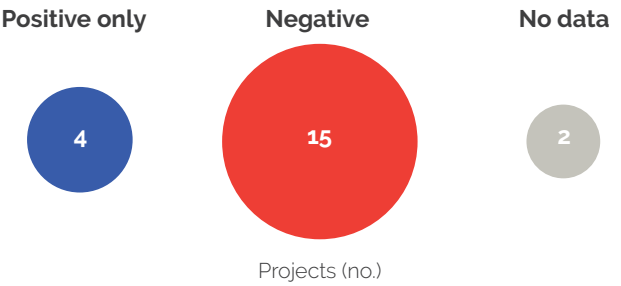
Note: Negative = the identified factor was reported as a constraint to project implementation. Positive = the identified factor was reported as facilitating implementation. Both = at the project level, there were positive and negative factors in the same category. This is more prominent in categories that were not disaggregated, such as coordination and engagement. For example, the Implementation Completion and Results Report showed that there was a clear allocation of roles and responsibilities (positive), but the bureaucratic structure created challenges to project implementation (negative). FY = fiscal year; *RAP* = *Results and Performance of the World Bank Group*.

Table 2.1. Risks Insufficiently Identified and Mitigated

Risk Types	Projects (*n* = 56; %)
Implementation capacity	38
Not specified	16
Political	13
Fiduciary	7
Environmental	5
Governance	5
Safeguards	5
Operational	4
Legislation	2
Economic	2
Stakeholders	2
Market response	2

Source: Independent Evaluation Group.

Figure 2.7. Inadequate Risk Identification and Mitigation of Weak Institutional Capacity and Low Technical Capacity of Implementing Agencies



Source: Independent Evaluation Group.

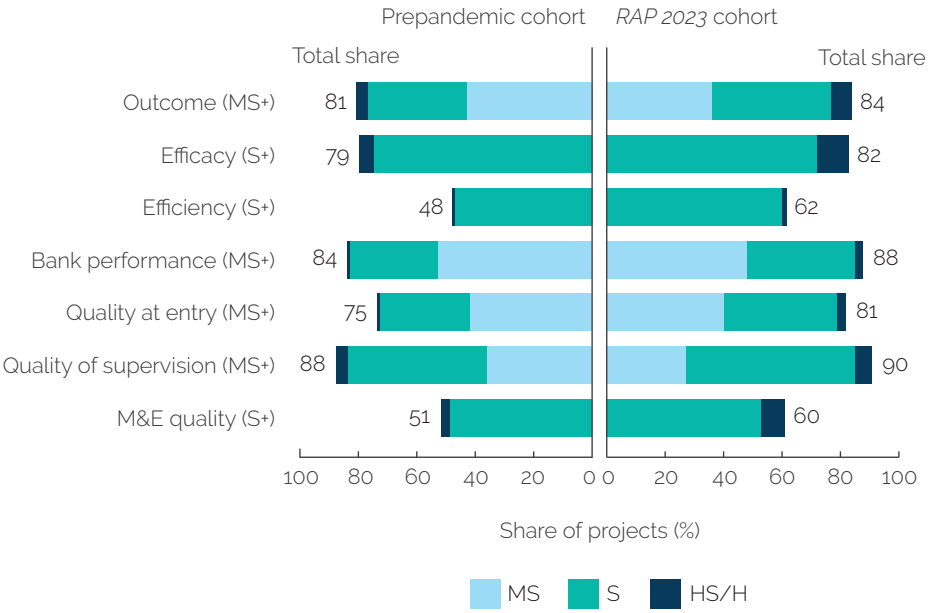
Note: The figure shows that the majority of projects that failed to adequately identify and mitigate capacity risks also reported the low technical capacity of implementing agencies as a challenge for implementation. Positive only = skilled human and organizational capacity was reported as facilitating project implementation. Negative = skilled human and organizational capacity was reported as a constraint to project implementation. No data = skilled human and organizational capacity issues were not reported by the project.

Factors that affected project implementation in the prepandemic cohort had a more adverse impact in the *RAP 2023* cohort. This *RAP*'s machine learning exercise, which expanded the analysis of factors to the prepandemic cohort, revealed that a larger share of projects experienced obstacles during implementation compared with previous years (figure 2.6). Among contextual factors, implementation challenges linked to conflict and instability increased from 8 percent of projects in the prepandemic cohort to 27 percent in the *RAP 2023* cohort. Similarly, natural disasters negatively affected the implementation of 23 percent of projects compared with 9 percent in previous years. Among stakeholders' dynamic factors, coordination and engagement challenges increased from 8 percent of projects in the prepandemic cohort to 26 percent in the *RAP 2023* cohort. Challenges caused by commitment and leadership changes among stakeholders undermined a larger share of projects than in previous years, increasing from 9 percent to 27 percent. Project finance–related challenges, particularly procurement, were also more frequently reported in the *RAP 2023* cohort (31 percent of projects) compared with the prepandemic cohort (15 percent of projects). It is important to highlight that these challenges cannot be fully attributed to the COVID-19 crisis. This is because the implementation phase of *RAP 2023* projects goes all the way back to 2003, making it impossible to determine if specific factors occurred during the pandemic or before. Furthermore, previous studies have identified similar challenges to project implementation, indicating that these are not unique to the COVID-19 pandemic.[7]

Project performance remained resilient to these implementation challenges. Overall, projects in the *RAP 2023* cohort performed better than those in the prepandemic cohort across all project ratings (figure 2.8).[8] Moreover, the World Bank's efficacy ratings in pursuing intended development outcomes have consistently improved in the long run (box 2.2). Only a few factors that affected implementation were statistically associated with project performance ratings, and their influence was moderate (figure 2.9). For example, 65 percent of projects that reported skilled human resources and organizational capacity as critical factors had an average outcome rating of 4.3, which is moderately satisfactory, compared with 4.6, which is satisfactory, for projects that did not report such issues. Previous studies, including

those by Denizer, Kaufmann, and Kraay (2013) and Ortega Nieto, Hagh, and Agarwal (2022), have also identified the negative association between human and organizational capacity weaknesses in both project outcomes and Bank performance. In addition, projects that identified key risks during the project preparation phase and outlined mitigation measures for them had an average outcome rating of 4.5 compared with 4.3 for projects that did not identify such risks (see box D.1 for other factors exhibiting a mild association with project performance ratings).

Figure 2.8. World Bank Project Ratings: The Prepandemic Cohort Compared with the *RAP 2023* Cohort



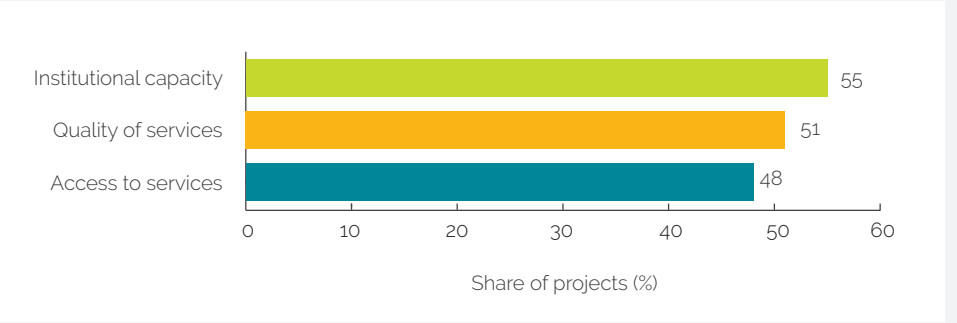
Source: Independent Evaluation Group.

Note: H = high; HS = highly satisfactory; M&E = monitoring and evaluation; MS = moderately satisfactory; MS+ = moderately satisfactory or above; *RAP* = *Results and Performance of the World Bank Group*; S = substantial or satisfactory; S+ = substantial or above (satisfactory or above).

Box 2.2. Development Outcomes Underlying Efficacy Ratings

Ratings increases in fiscal years (FY)20–22 are not a result of a systematic difference in the projects' intended development outcomes compared with previous years. The analysis of outcome types indicates that the top three development outcomes pursued by the World Bank, as observed in the *Results and Performance of the World Bank Group 2021* across FY12–14 and FY17–20 (second quarter), continue to be increasing institutional capacity, improving service quality, and expanding access to services (figure B2.2.1; see appendix A for methodology and appendix C for more details).

Figure B2.2.1. Top Three Development Outcomes in the *RAP 2023* Cohort



Source: Independent Evaluation Group.

The efficacy ratings have shown a consistent improvement over time, and this upward shift is statistically significant in the long run. The comparison between FY12–14 and FY17–20 (second quarter), as well as between FY12–14 and FY20–22, demonstrates statistically significant improvement. This indicates the World Bank's ongoing efforts to enhance project efficacy and effectiveness, which are reflected in the improved performance ratings observed in FY20–22.

(continued)

Box 2.2. Development Outcomes Underlying Efficacy Ratings (cont.)

Table B2.2.1. Average Efficacy Rating by Objective Outcome Type

Outcome Type	Percentage of Objectives			Average Efficacy Rating			Statistical Significance in Difference in Average Efficacy Rating		
	FY12–14	FY17–FY20 (Q2)	FY20 (March) – FY22	FY12–14	FY17–FY20 (Q2)	FY20 (March)– FY22	FY12–14 vs. FY17–FY20	FY17–FY20 (Q2) vs. FY20 (March)– FY22	FY12–14 vs. FY20 (March)– FY22
Capacity of institutions enhanced	37	40	33	2.43	2.72	2.70	Yes	No	Yes
Quality of services improved	40	47	36	2.59	2.77	2.83	Yes	No	Yes
Access to services expanded	23	25	30	2.63	2.85	2.87	Yes	No	Yes

Source: Independent Evaluation Group.

Source: Independent Evaluation Group.
Note: The periods of FY12–14 and FY17–20 (second quarter) include only a sample of projects that represent 29 percent and 31 percent of the population, respectively. FY = fiscal year; Q = quarter.

Figure 2.9. Relationship between Implementation Factors and Project Ratings in *RAP 2023* Cohort

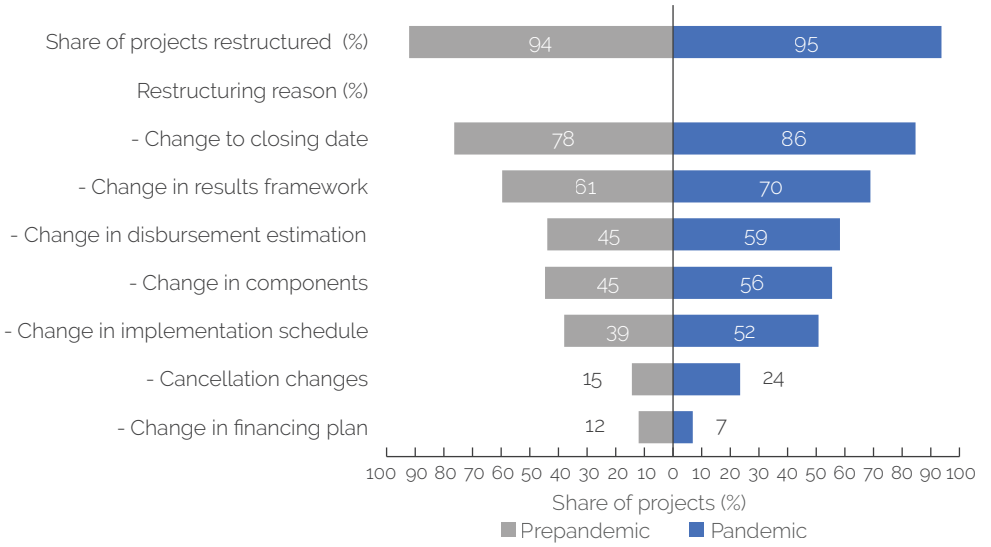
Source: Independent Evaluation Group.

Note: Differences in average ratings between projects that identified the implementation factor and those that did not were statistically significant, as determined by both *t*-test and Mann-Whitney *U* test. M&E = monitoring and evaluation; *RAP* = *Results and Performance of the World Bank Group*.

Adaptation and Restructuring for Results

More project adaptation and restructuring during implementation may explain the improved project performance. The adaptive and learning capacity of project teams enabled them to overcome implementation challenges (which helps explain the limited impact of these challenges on project performance). *RAP 2020* anticipated that projects would require more frequent course corrections to adapt and respond to unexpected shocks, including those related to the pandemic (World Bank 2020). Indeed, IEG's evaluation on the World Bank's early response to COVID-19 showed that repurposing existing projects allowed the World Bank to rapidly adapt to the pandemic (World Bank 2022b).[9] This *RAP*'s analysis confirms that there was a notable change in project restructuring patterns during the pandemic (figure 2.10). An examination of restructuring dates revealed that restructurings occurred more frequently after March 2020, which coincides with the onset of the pandemic (table D.2). Overall, the number of restructurings increased from an average of 1.9 per project in the prepandemic cohort to 2.6 in the *RAP 2023* cohort.

Figure 2.10. Occurrence and Reasons for Restructuring: The Prepandemic Cohort Compared with the *RAP 2023* Cohort



Source: Independent Evaluation Group.

Note: RAP = Results and Performance of the World Bank Group.

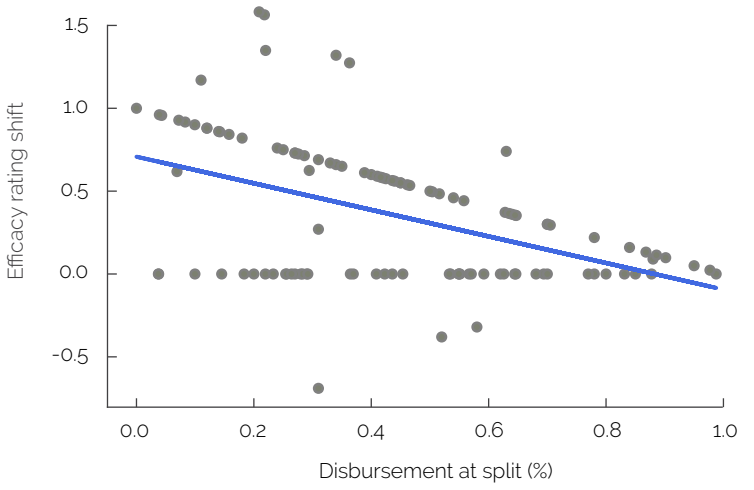
Extensions of project closing dates seem to have helped projects achieve their intended outcomes. As expected from the delays caused by the pandemic, the share of projects that changed their closing dates increased from 78 percent in the prepandemic cohort to 86 percent in the *RAP 2023* cohort. Among these, project extensions accounted for 79 percent of these changes and accelerated closing dates accounted for the rest.[10] Project extensions compensated for the time lost during lockdowns, likely providing the time needed to achieve their intended development outcomes. At the same time, project extensions can increase costs, with implications for projects' cost-effectiveness and cost benefit. However, project extensions did not affect project efficiency ratings. The share of projects rated with substantial or high efficiency was higher in the *RAP 2023* cohort (62 percent) than in the prepandemic cohort (48 percent), and the correlations between project extensions and average efficiency ratings were not statistically significant.[11]

Changes in results frameworks helped projects achieve their intended outcomes. Restructuring data show a notable increase in the share of projects that revised their results frameworks, rising from 61 percent in the prepandemic cohort to 70 percent in the *RAP 2023* cohort. These changes in results frameworks entailed replacing indicators for better measurements; adding new indicators to account for changes to a project's scope, for example, in a project that expands into a new geographical area; and changing indicator targets to respond to unexpected changes in the project's circumstances, such as changes caused by the pandemic or made because targets at project appraisal were no longer, or had ever been, realistic, among other changes. An in-depth review of a sample of 54 ICRRs with modest M&E quality shows that revising results frameworks during implementation helped these modestly rated projects improve to have substantial efficacy ratings. In the sample, 93 percent had shortcomings in the initial design of their results frameworks (table C.9). These shortcomings included (i) inadequate selection of indicators, (ii) a lack of a data collection methodology, (iii) unrealistic targets, and (iv) attribution issues. However, many project teams were able to rectify these shortcomings during implementation by refining their M&E methodology, revising indicators, or adjusting targets through project restructuring. Gathering additional evidence on projects' achievements to supplement results frameworks—such as qualitative information, impact

evaluation findings, or beneficiary survey data—resulted in a good efficacy rating of substantial (see appendix C). That said, more analysis is needed on the type of revisions that project teams made to project results frameworks to understand how specific changes influence efficacy.[12]

Timely course corrections to results frameworks also helped projects achieve their intended development outcomes. Some of these restructurings led IEG to adopt a split rating methodology, which can occur when both (i) teams revise project objectives or associated outcome targets during implementation and (ii) project achievements of original objectives or targets differ from revised objectives or targets.[13] Indeed, the share of projects with split ratings increased from 3 percent in the prepandemic cohort to 22 percent in the *RAP 2023* cohort. In addition, the evidence shows that the earlier these revisions occur in the project cycle, the greater the likelihood that projects will achieve their intended development outcomes. Figure 2.11 shows that the earlier revisions occur during the project life, the higher the project's efficacy ratings are compared with what they would have been without the revision of the original objectives or key associated outcome targets (see also table D.3).

Figure 2.11. Timing of Project Revisions and the Shift in Efficacy Rating in Fiscal Years 2019–22 When a Split Rating Is Applied



Source: Independent Evaluation Group.

Note: The percentage disbursed at the split indicates the timing of project revision. The shift in efficacy rating refers to the difference between the final efficacy rating and the efficacy rating applied when considering the original target. The blue line is a regression line showing the correlation between disbursement at split (%) and the shift in efficacy rating.

Monitoring and Evaluation for Adaptation and Results

Improvements in M&E quality facilitated project adaptation and helped provide sufficient evidence on projects' achievements. Several studies describe M&E as an early warning mechanism that enables effective adaptive management (Denizer, Kaufmann, and Kraay 2013; Ika, Diallo, and Thuillier 2012; World Bank 2016, 2020, 2021). Strong M&E frameworks equip teams with a deep understanding of project challenges, allowing them to address weaknesses, make timely course corrections, and achieve desired development outcomes. Previous evidence shows that World Bank projects with strong M&E frameworks have higher (and statistically significant) outcome ratings (Raimondo 2016; World Bank 2020, 2021). Similarly, projects in the *RAP 2023* cohort with higher M&E quality ratings had higher efficacy ratings (table 2.2). This is not surprising because efficacy ratings take into account both the validity of the results framework to measure the intended development outcomes and the actual achievement of those outcome measures. Furthermore, ICRR data indicate that projects with modest or negligible efficacy ratings mostly failed to achieve well-defined target indicators, or had low achievement, rather than failed to define appropriate results framework indicators, or had insufficient evidence (figure 2.12).[14]

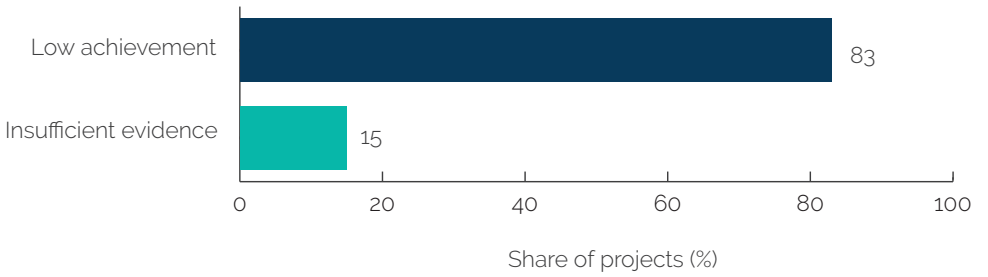
Table 2.2. Overall Efficacy and Monitoring and Evaluation Quality Ratings (percentage of projects)

	Overall Efficacy			
M&E Quality	Negligible	Modest	Substantial	High
Low	0.4	1.4	0.0	0.0
Modest	1.4	13.0	23.0	0.0
Substantial	0.0	1.4	45.0	7.0
High	0.0	0.0	4.0	4.0

Source: Independent Evaluation Group.

Note: M&E = monitoring and evaluation.

Figure 2.12. Reason for Low Efficacy among Projects Rated Negligible or Modest



Source: Independent Evaluation Group.

Note: In the *Results and Performance of the World Bank Group 2023* cohort, only 15.4 percent of projects received ratings of negligible or modest efficacy.

Results frameworks with well-aligned and adequate indicators contributed to improved efficacy ratings. In assessing the validity of results frameworks in measuring intended development outcomes and their associations with efficacy ratings, we examined 4,808 indicators corresponding to the 273 projects included in the *RAP 2023* cohort. Indicators were classified according to their (i) outcome type; (ii) adequacy—fully, partially, or not adequate—in accurately measuring the individual objectives;[15] and (iii) level (including output, intermediate outcome, outcome, or high outcome; see more details on these classifications in appendixes A and C). Our analysis found strong alignment between indicators and outcome types, with 97 percent of objectives having indicators of the same outcome type; moreover, objectives with indicators well aligned to them tend to have higher efficacy ratings (see table C.10). The analysis also found that 85 percent of development objectives had at least one fully adequate indicator to measure a project's intended development outcome. On average, individual objectives had 65 percent fully adequate, 35 percent partially adequate, and 0 percent inadequate PDO indicators.[16] The adequacy of indicators also matters for both objective efficacy and overall project efficacy. Objectives with more fully adequate indicators tended to have higher efficacy ratings.

However, the outcome orientation of results frameworks does not explain efficacy ratings. Our analysis found that most results framework indicators

were at the outcome and intermediate outcome levels. Forty percent of PDO indicators that measured the achievement of individual project objectives measured outcomes, 46 percent measured intermediate outcomes, 12 percent measured outputs, and a mere 2 percent measured high outcomes. Most intermediate results indicators for projects were even less outcome oriented because they were mainly lower-level indicators designed to gain insights on the project's progress toward completing project activities. These indicators mostly measured outputs (54.0 percent), followed by intermediate outcomes (38.0 percent), outcomes (8.0 percent), and high outcomes (0.2 percent). This *RAP*, however, found no significant associations between a project's indicator level and its efficacy ratings. One explanation for this is that objectives without outcome-level indicators may still yield a substantial efficacy rating provided that other lower-level indicators demonstrate that the project completed the intended activities and that these activities would plausibly contribute to the achievement of intended development outcomes, as outlined by the project's theory of change. Another explanation for the lack of correlation is related to the nature of intended development objectives. Development objectives that aim for intermediate outcomes do not need high outcome indicators to measure achievement. This is typically the case for development objectives that aim to increase access to services (box 2.3).

Box 2.3. Development Outcomes Underlying Efficacy Ratings and Validity of Results Frameworks

Among the 16 types of development outcomes classified, the World Bank has been more successful in expanding access to services than in improving quality of services or enhancing institutional capacity.

Expanding access to services was the intended development outcome with the highest efficacy rating (average of 3.1 on a 4-point scale, which is substantial). Objectives pursuing this type of outcome also outperformed others in the adequacy of indicators. Many objectives aiming at expanding access to services were stated as lower-level results, thus not requiring high outcome indicators to measure and demonstrate achievements. On average, 74 percent of project development objective indicators measured outputs and intermediate outcomes, whereas only 26 percent measured outcomes.

(continued)

Box 2.3. Development Outcomes Underlying Efficacy Ratings and Validity of Results Frameworks (cont.)

Objectives aiming at improving the quality of services also had on average a substantial efficacy (average of 3.0 on a 4-point scale), along with a high adequacy of indicators. Indicators measuring the quality of services were found at all four levels, in line with the specific dimensions of quality that the project focused on. For example, objectives addressing quality of services had on average 15 percent of indicators at output level that measured improvements in structural quality, such as rehabilitating or upgrading infrastructure and training service providers. They had 48 percent of outcome-level indicators that measured, for example, time savings and user satisfaction with services provided and only 5 percent of high outcome–level indicators that measured, for example, fatality rates.

Enhancing the capacity of institutions to perform remains a particularly challenging outcome to achieve. Objectives targeting this outcome type received statistically significant lower efficacy ratings, with an average of 2.8 on a 4-point scale. Consistent with the findings of *Results and Performance of the World Bank Group 2021*, the attainment of these development outcomes was measured predominantly by intermediate outcome or lower-level indicators (67 percent). For further details, see appendix C.

Table B2.3.1 Individual Objective Efficacy, Level, and Adequacy of Results Framework Indicators

Top Three Development Outcomes	Average Individual Objective Efficacy Rating (4-point scale)	PDO Indicator-Level Score (4-point scale) and Share of PDO Indicators by Level (%)	PDO Indicator Adequacy Score (4-point scale)
Access to services expanded	3.06 (*β*)	2.20 (*α*) High outcome 0 Outcome 26 Intermediate outcome 68 Output 6	2.78 (*β*)

(continued)

Box 2.3. Development Outcomes Underlying Efficacy Ratings and Validity of Results Frameworks (cont.)

Top Three Development Outcomes	Average Individual Objective Efficacy Rating (4-point scale)	PDO Indicator-Level Score (4-point scale) and Share of PDO Indicators by Level (%)	PDO Indicator Adequacy Score (4-point scale)
Quality of services improved	2.95	2.43 (*α*,*γ*) High outcome 5 Outcome 48 Intermediate outcome 32 Output 15	2.70 (*γ*)
Capacity of institutions to perform institutional functions enhanced	2.76 (*ß*)	2.17 (*γ*) High outcome 0 Outcome 33 Intermediate outcome 50 Output 17	2.46 (ß,*γ*)

Source: Independent Evaluation Group.
Note: Statistical significance at least 0.05 based on Student *t*-test and Mann-Whitney *U* test. *α* = statistically significant difference between access to services expanded and quality of services improved; *ß* = statistically significant difference between access to services expanded and capacity of institutions to perform institutional functions enhanced; PDO = project development objective; *γ* = statistically significant between quality of services improved and capacity of institutions to perform institutional functions enhanced.

Source: Independent Evaluation Group.

[1] A project's exposure time to COVID-19 was calculated as the period from March 2020 until the project's closing date divided by the project's overall duration. For a small number of still-active projects that had an Implementation Completion and Results Report Review (ICRR) completed, the exposure measure was calculated as the period from March 2020 until the project's ICRR completion date divided by the project's overall duration.

[2] These two major overlapping shocks to the global economy over the past three years had a significant impact on economic growth across regions, by stoking uncertainty and disrupting global trade and supply chains. The resulting increases in energy, food, and fertilizer prices also amplified the inflationary pressures (World Bank 2023a).

[3] The evaluation methodology for development policy financing projects changed in mid-2020. In the old methodology, Bank performance overall rating was based on quality at entry and quality of supervision, whereas in the new methodology, it is based on design and implementation (see appendix A).

[4] In addition to the COVID-19 pandemic, seven projects also reported encountering multiple concurrent outbreaks, including Ebola, cholera, and measles. However, the COVID-19 pandemic emerged as the most frequently cited among them.

[5] The content analysis of self-reported factors affecting implementation of this *Results and Performance of the World Bank Group* (*RAP*) identified both challenges and enablers faced by projects, as stated in the Implementation Completion and Results Report narrative, specifically focusing on the Factors Affecting Project Implementation and Performance section. For factor classification purposes, an adapted version of the Delivery Challenges in Operations for Development Effectiveness taxonomy was used (see appendix A for methodology and appendix D for more details on factors that affected implementation).

[6] See figures D.4 through D.9 for details on the distribution of factors affecting implementation by project subgroups.

[7] Ortega Nieto, Hagh, and Agarwal (2022) used data from the Delivery Challenges in Operations for Development Effectiveness developed by the Global Delivery Initiative. Their study examined project performance and the attainment of development objectives across 42 specific delivery challenges, drawing from a data set of over 5,000 lending projects spanning the period from 1995 to 2015.

[8] Performance rating improvements between the prepandemic and the *RAP 2023* cohorts are not attributed to a systematic difference in the composition of the portfolio. The decompo-

sition analysis shows that the primary factor contributing to the overall increase in performance ratings is not portfolio changes but rather rating increases within various subgroups (including Global Practice, Region, project size, country income level, lending group, and fragile and conflict-affected situation status; see figures C.11 through C.15).

[9] Approximately 60 percent of World Bank country programs underwent a significant reorientation of their portfolios to adapt to the changing needs caused by COVID-19, involving extensive repurposing of projects, additional support through advisory services and analytics, and the introduction of new initiatives.

[10] However, this does not mean that project duration became longer. There was no statistically significant difference in the length of the extensions because both cohorts had a mean and median extension period of 15 months and 12 months, respectively (see figure D.14).

[11] The December 2020 guidance note, developed by Operations Policy and Country Services in collaboration with the Independent Evaluation Group, titled "Preparing an ICR for a Project Impacted by COVID-19," acknowledges that certain delays in project implementation may not necessarily indicate inefficiency, particularly if there was an ongoing active response. This helps to explain the lack of correlation between project extensions and efficiency ratings.

[12] An in-depth analysis of project restructurings papers that described the changes made in results frameworks and the assessment of the extent to which targets set at the outset were realistic are outside the scope of this *RAP*. A preliminary review of available guidelines on setting indicators' targets, conducted at the Concept Note stage, suggests that the assessment of the adequacy of target levels needs to consider the historical trends of that particular indicator, benchmarking (results achieved by similar projects), expert judgment, and stakeholder expectations. Moreover, setting targets will depend on context-specific factors such as available resources, institutional capacity, environmental and political concerns, the duration of the project, the complexity of the intervention, and the contribution of other donors' inputs. Such a detailed assessment cannot be realistically undertaken at scale by the *RAP* product, which uses other Independent Evaluation Group micro products as the main sources for evidence.

[13] According to ICRR guidelines, Independent Evaluation Group staff independently assess the appropriateness of applying a split rating versus assessing the entire project. A split rating typically applies when (i) the project objectives or key associated outcome targets were revised during implementation and (ii) the project's achievements based on original objectives and targets differed from those based on revised objectives and targets. For example, if the project expanded its scope, and the targets for the original geographical areas were achieved, but the targets for the new geographical areas added at restructuring were not achieved, then

a split rating is applied. When the project's scope has decreased through a downward revision of targets, and the original target was not achieved, but the revised target was achieved, a split rating is also applied. When deriving the project's overall efficacy and outcome ratings, the split rating takes into account the project's achievements against both the original and the revised objectives and targets, weighted by the disbursement rate at the time of the revisions. See section 9 of *Guidelines for Reviewing World Bank Implementation Completion and Results Reports: A Manual for IEG ICR Reviewers* (World Bank 2017).

[14] Since the introduction of the reasons for including a low efficacy rating in the ICRR system in 2017, there have been no significant changes in the share of unsuccessful projects attributed to either of these reasons over time (see figure C.16).

[15] For example, in a Transport project, the indicators of reopened project roads in good to fair condition and roads in good and fair condition as a share of total classified roads were fully adequate because they can demonstrate the achievement of the individual objective to reestablish lasting road access between provincial capitals, districts, and territories in the project impact area. The indicator of number of condoms distributed, instead, was not adequate because it did not provide evidence toward the improvement of roads conditions, and the indicator of action plan to develop the road construction industry implemented was partially adequate because it contributed to demonstrating the achievement of the individual objective to some extent, but it is not sufficient (see appendix A for methodology and appendix C for more details on the analysis).

[16] When considering all indicators included in project results frameworks (that is, project development objective and intermediate results indicators), individual objectives have on average 30 percent of fully adequate indicators, 69 percent of partially adequate indicators, and just 0.4 percent of not adequate indicators.

3 International Finance Corporation Results and Performance

International Finance Corporation (IFC) investment projects included in the calendar year (CY)20–22 *Results and Performance of the World Bank Group* cohort were moderately or minimally affected by COVID-19 (as assessed by IFC at the time of sampling). These investment projects were exposed to the pandemic for 24 percent of their project lives. As such, it is too early to assess the full impact of COVID-19 on IFC investment projects. That said, COVID-19–related lockdowns and economic slowdowns contributed to a more challenging operating environment for the CY20–22 cohort.

Along with COVID-19, IFC investment projects were negatively affected by unfavorable economic issues, high business risks, and high competition. IFC has no formal procedures for modifying the original development objectives, indicators, and targets to adapt to changing market conditions.

Notwithstanding the challenging environment and the inability of IFC to restructure project objectives and targets, IFC's investment project development outcome success ratings declined only slightly from 53 percent in CY19–21 to 50 percent in CY20–22. IFC's Expanded Project Supervision Report self-ratings also showed a decline.

Private sector sponsors and clients reacted quickly to the changing economic landscape during the pandemic. The strong ability and technical expertise of sponsors contributed to adaptive management and were the factor that most positively affected investment project performance.

This *Results and Performance of the World Bank Group* confirms that IFC investment project objectives were highly outcome oriented, with all projects pursuing project-level outcomes and a majority (74 percent) also pursuing market-level outcomes beyond the project. However, IFC investment projects' outcome achievement rates were relatively low, and a lack of appropriate results indicators and evidence constrained some outcome measurements. Investment projects with high outcome achievement rates had higher development outcome ratings.

IFC advisory projects' development effectiveness ratings slightly declined from 60 percent in the fiscal years (FY)19–21 to 54 percent in FY20–22. IFC's Project Completion Report self-ratings also showed a decline. The challenging operating environment and weaknesses in advisory project preparation and monitoring and evaluation contributed to the ratings slide.

This chapter presents trends and patterns of IFC's investment and advisory project performance.[1,2] It describes the context of IFC's operating environment, including how the markets test IFC private sector projects' efficiency and competitiveness. The chapter also explores IFC investment projects' development outcomes and key factors influencing project implementation and performance.

Project Exposure to the COVID-19 Pandemic and Sample Selection Bias

It is too early to assess the full impact of COVID-19 on IFC investment projects. IFC's *RAP* cohort includes investment projects that were approved during CY12–17 and evaluated in CY20–22, when they achieved their early operating maturity stages;[3] 75 percent of IFC investment projects in the *RAP* cohort were still active. The cohort projects' average exposure to COVID-19 was 24 percent of their active project lives—a larger percentage than for the World Bank's *RAP* cohort. However, IFC's *RAP* cohort did not include investment projects that were severely affected by the pandemic because IEG agreed that IFC could defer the project evaluations for these projects.[4,5] These changes in the sampling processes influenced the profile of CY20–22 investment projects, creating a sample selection bias in which only projects moderately or minimally affected by COVID-19 were included. As such, the *RAP*'s analysis provides only preliminary insights into how the pandemic affected IFC investment project implementation and performance.

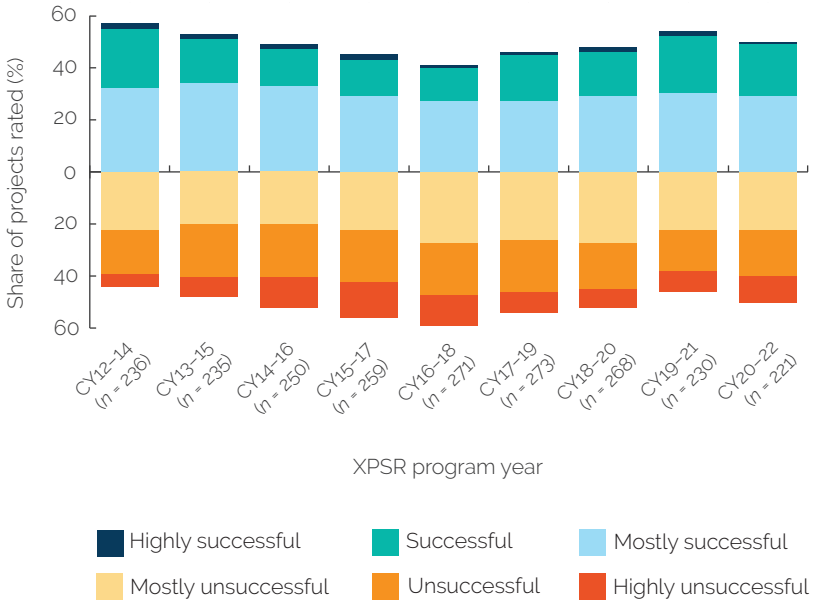
Preliminary findings suggest that COVID-19 undermined the implementation of CY20–22 investment projects. COVID-19 caused lockdowns, supply chain disruptions, asset quality issues, and an economic slowdown—all of which affected investment project implementation. The lockdowns particularly affected investment projects in the real sector, shutting down or limiting the operations of hotels, hospitals, transportation companies, manufacturing facilities, and tertiary education providers. In addition, the lockdowns, combined with the overall economic downturn, led to reduced demand for products and services in most sectors and for most IFC clients.

IFC has no formal procedures for modifying investment projects' development objectives to adapt to changing market conditions subsequent to the Board approval of a project. By their nature, private sector projects must be financially sustainable and survive in a competitive market to be viable. Moreover, all IFC investment projects are also required to comply with IFC's environmental and social performance standards. If needed, IFC can restructure the terms of investment financing agreements with clients and reschedule loan repayment schedules, and clients can adapt their products and services to changing market conditions, such as those caused by the COVID-19 pandemic. However, the original development objectives, indicators, and targets cannot be changed to reflect the changes in market conditions since neither IFC processes nor the Anticipated Impact Measurement and Monitoring (AIMM) framework consider formal changes of development objectives or targets after Board approval.

Project Performance Rating Trends

Development outcome success ratings for IFC investment projects declined only slightly in CY20–22 despite the difficult operating environment. The share of IFC investment projects with outcomes rated mostly successful or better had been increasing since CY16–18 but declined slightly from 53 percent in CY19–21 to 50 percent in CY20–22. IFC's XPSR self-ratings also showed a decline. The decline was driven by the lower performance of CY22 investment projects, where the success rates dropped from 59 percent in CY21 to 50 percent in CY22. This decline in development outcome ratings mainly reflected lower average ratings across most project subgroups. Changes in the evaluated portfolio shares of different project subgroups did not have a significant effect on the overall decline (see the decomposition analysis in appendix E). On a granular basis, the share of projects with unsuccessful and highly unsuccessful ratings increased, the share of projects with mostly unsuccessful ratings remained the same, and the share of projects with mostly successful, successful, or highly successful ratings shrank (see figure 3.1).

Figure 3.1. IFC Investment Project Development Outcome Ratings



Source: Independent Evaluation Group, XPSR database.

Note: CY = calendar year; IFC = International Finance Corporation; XPSR = Expanded Project Supervision Report.

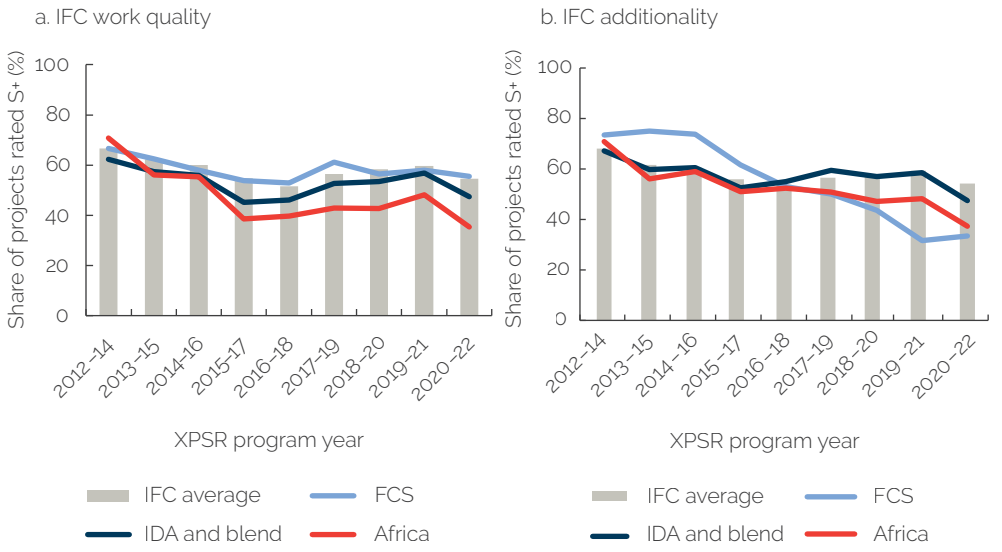
The development outcome ratings of investment projects in African, FCS, and IDA and blend countries were substantially low. The shares of African, FCS, and IDA and blend investment projects rated mostly successful or better for development outcome were 27 percent, 11 percent, and 36 percent in CY20–22, respectively. This was mainly driven by adverse macroeconomic factors, the pandemic's effects, high business risks, and low sponsor or client management quality. Continued low performance in these markets would undermine the IFC 3.0 strategy, which aims to address conflict, fragility, and forced displacement by increasing operations in FCS and IDA and blend countries. Indeed, the share of investments in IDA and blend countries in IFC's overall portfolio increased from 27 percent to 32 percent between 2019 and 2022, whereas the share of investments in FCS countries grew from 7 percent to 10 percent during the same period.

IFC investments, unlike World Bank projects in the public sector, must overcome hurdles unique to the private sector to be successful. IFC is a minority investor in projects alongside private sector sponsors and, therefore, shares risks, including commercial risks, with these other investors. In FCS and

IDA and blend countries, IFC investment projects are often constrained by the limited number of potential sponsors with adequate capacity, resources, and relevant experience to undertake such investment projects. The private sector often faces difficult investment climates and regulatory environments in such countries. As such, investment projects could continue to deteriorate without up-front and upstream efforts to improve the business environment, reduce investment risks, attract private investors, and build sponsor capacity in these countries.

IFC work quality, which is important for overcoming these hurdles, weakened in CY20–22. Overall, the share of IFC investment projects with satisfactory or better IFC work quality ratings was 55 percent in CY20–22 (down from 60 percent in CY19–21).[6] The share of investment projects with high project preparation work quality ratings decreased from 59 percent to 54 percent between CY19–21 and CY20–22, whereas high supervision work quality ratings stayed at approximately 70 percent during the same period. The decline in IFC work quality ratings was more pronounced in projects in African and IDA and blend countries compared with other country types (figure 3.2, panel a). IFC work quality ratings in FCS investment projects were on par with those of non-FCS investment projects. The challenging environment in African, IDA and blend, and FCS countries compels IFC to conduct more thorough due diligence, risk mitigation, and investment structuring at project preparation and provide enhanced implementation support during supervision. As in previous *RAP*s, *RAP 2023* confirmed that IFC work quality ratings, particularly for project preparation, are positively and strongly associated with development outcome ratings.[7] For example, *RAP 2022* stated that there is a strong association between IFC work quality both at the front end and at implementation and the development outcome ratings (World Bank 2022b, 17). *RAP 2021* and *RAP 2020* both had similar findings.

Figure 3.2. IFC Investment Project Work Quality and Additionality Ratings



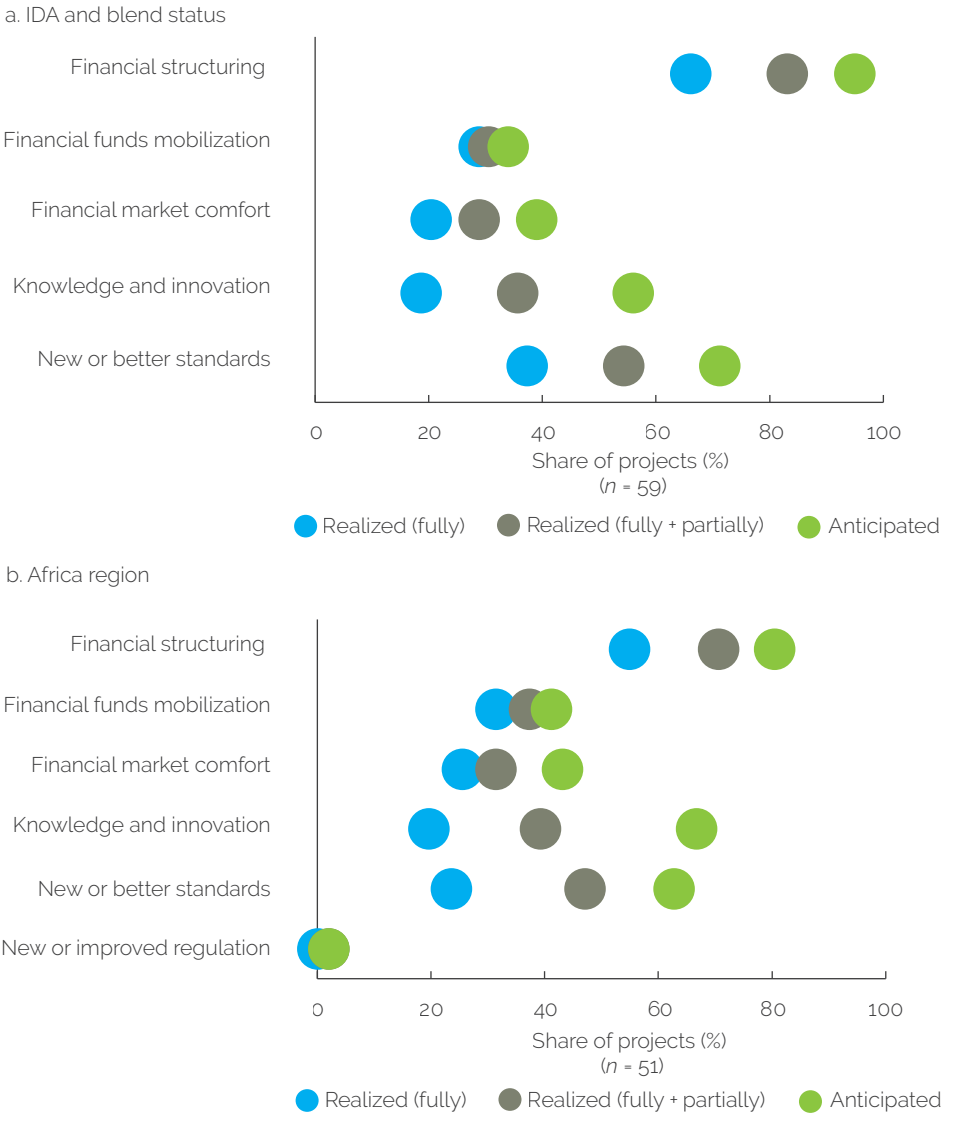
Source: Independent Evaluation Group, XPSR database.

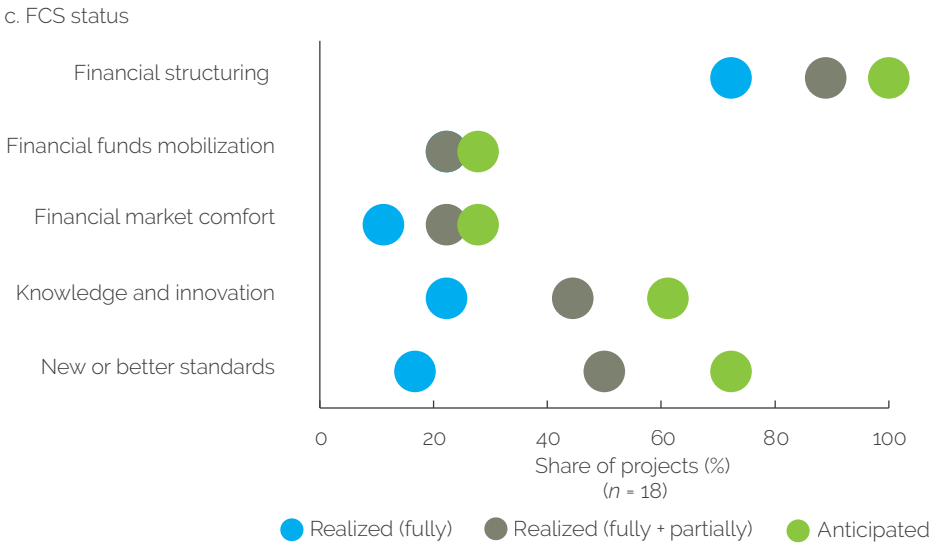
Note: FCS = fragile and conflict-affected situation; IDA = International Development Association; IFC = International Finance Corporation; S+ = satisfactory or better; XPSR = Expanded Project Supervision Report.

IFC additionality success ratings in challenging environments were lower than the IFC average. Overall, the share of IFC investment projects with high additionality ratings was 54 percent in CY20–22 (down from 59 percent in CY19–21). In challenging environments, such as FCS, African, and IDA and blend countries, IFC realized its anticipated additionality in 33 percent, 37 percent, and 47 percent of projects, respectively (figure 3.2, panel b). The gap between anticipated and realized additionality in these challenging markets was larger for nonfinancial additionality than for financial additionality. For example, the gap in provision of knowledge and innovation additionality was 27 percent in African, 20 percent in IDA and blend, and 17 percent in FCS countries.[8] The gap in setting new or better standards, for example, in environmental and social and corporate governance practices was 16 percent in African, 17 percent in IDA and blend, and 22 percent in FCS countries (figure 3.3). IEG's recent evaluation on IFC additionality in middle-income countries also found that IFC had the most difficulty with delivering nonfinancial additionality (World Bank 2023b), which requires more proactive supervision and implementation during an investment project's life cycle. Within the *RAP 2023* cohort, IFC additionality ratings

were positively and strongly correlated with development outcome ratings.[9] *RAP 2022* also noted that IFC additionality was closely associated with development outcome ratings and that "IFC considers additionality essential to achieving development impact" (World Bank 2022a, xv).

Figure 3.3. Comparison of Anticipated and Realized Additionalities for Select Project Categories





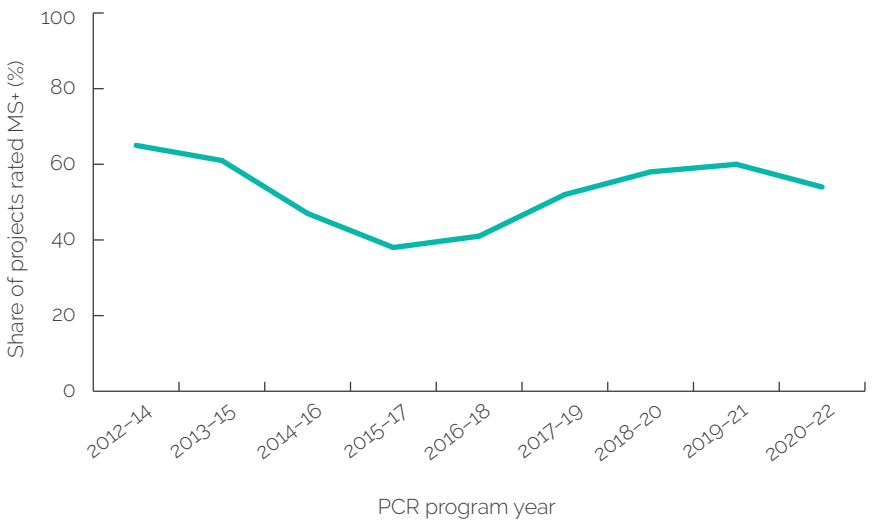
Source: Independent Evaluation Group.

Note: The financial structuring, financial funds mobilization, and financial market comfort are the types of financial additionality. Conversely, knowledge and innovation, new or better standards, and new or improved regulation are the types of nonfinancial additionality. FCS = fragile and conflict-affected situation; IDA = International Development Association.

Overall, IFC investment outcome success ratings declined, although its equity performance remained stable. Financial sustainability is important for individual IFC project success and for IFC's own sustainability as an investor and institution. IFC overall investment outcome ratings have been satisfactory or better in 60 percent of investment projects in CY20–22, which was slightly lower than 64 percent in CY19–21. This decline was caused by the slight decline in loan investment outcome ratings, some of which were caused by prepayments.[10] In contrast, equity outcome ratings have remained stable, although only about a third of equity investments generated satisfactory returns. IFC achieved a "double bottom line" of high development outcome ratings and high investment returns in 42 percent of investment projects. The achievement of a double bottom line was lower in African, FCS, and IDA and blend countries, where a significant share of investment projects delivered neither positive development results nor satisfactory investment returns. Both IFC's development and investment outcome ratings were low in 51 percent of African projects, 56 percent of FCS projects, and 39 percent of IDA and blend projects.

IFC advisory projects' development effectiveness ratings declined slightly in the more challenging operating environment. The development effectiveness of IFC's advisory projects has been improving since FY15–17, but the success ratings declined from 60 percent to 54 percent between FY19–21 and FY20–22 (figure 3.4). The overall ratings decline was mainly caused by a decrease in average ratings and not by changes in the evaluated portfolio's composition (see the decomposition analysis in appendix E). IFC's self-ratings in Project Completion Reports for advisory projects also declined. Fifty-four percent of advisory projects achieved satisfactory or better outcomes by the project's completion date, despite 86 percent of projects delivering their outputs to the clients as expected. About a third of advisory projects had weak strategic relevance, whereas close to half had efficiency shortcomings. However, 23 percent of advisory projects managed to achieve longer-term impacts by the time of project completion. This is a positive achievement because advisory projects are not expected to achieve impacts by completion.

Figure 3.4. IFC Advisory Project Development Effectiveness Success Ratings



Source: Independent Evaluation Group, PCR database.

Note: IFC = International Finance Corporation; MS+ = mostly successful or better; PCR = Project Completion Report.

IFC's advisory project development effectiveness ratings varied across regions and primary business areas. The decline in performance was more

pronounced in the Financial Institutions Group; Public-Private Partnership; Manufacturing, Agribusiness, and Services; and Environment, Social, and Governance projects (which represented 61 percent of projects in IFC's portfolio). Advisory projects in these primary business areas had weaknesses in project preparation and design and in M&E. External factors such as political conflicts, force majeure events, COVID-19–related disruptions, and client commitment issues also negatively affected the more recent projects in this *RAP* cohort and contributed to their low development effectiveness ratings.

The development effectiveness ratings of advisory projects were highly correlated with IFC work quality ratings, particularly for project preparation and design work quality. The relationship between IFC work quality ratings and development effectiveness ratings in IFC advisory projects has been established in previous *RAP*s and the 2017 joint IEG-IFC work quality study. They were correlated in 79 percent of advisory projects in the *RAP* cohort. IFC overall work quality ratings were satisfactory in 59 percent of advisory projects in FY20–22. However, IFC's preparation and design work quality ratings were satisfactory or better in fewer than half of projects in FY20–22. The implementation and supervision work quality success ratings of advisory projects marginally declined in FY20–22, although 61 percent of these projects exhibited high work quality. Project design and preparation ratings were lower than the IFC average in advisory projects in the Africa region, with the success rate of 43 percent in FY20–22. Supervision and administration work quality success ratings continued to weaken in the Africa region and in IDA and blend countries. The share of African advisory projects with high implementation and supervision work quality ratings decreased from 58 percent in FY19–21 to 49 percent in FY20–22, while IDA and blend advisory projects saw a decline from 67 percent to 61 percent during the same period. Only 15 percent of the African advisory projects and 21 percent of IDA and blend projects with low IFC work quality achieved high development effectiveness ratings.

Factors Affecting Project Implementation and Performance

Several factors besides COVID-19 also negatively affected IFC's investment project performance. This *RAP* conducted a deep-dive analysis of 170 IFC investment projects from the *RAP* cohort to find common factors affecting

performance. The analysis identified the top three factors, among 5 categories and 51 subcategories, affecting performance for each project. These factors could have either a negative or positive influence on project performance (see appendix A for definition of different factors). We found that about a quarter of IFC investment projects in CY20–22 were negatively affected by unfavorable economic factors. These factors reduced demand for IFC client products and services and lowered the project companies' operational and financial performance compared with the projections at the Board approval stage. The second-most common negative factor was high business risks, which affected the performance of 17 percent of investment projects. Many financial sector projects moved away from lending to riskier segments, such as micro, small, and medium enterprises and affordable housing finance, because of the economic slowdown and increased credit risks. This risk management was needed to help preserve capital; however, the consequence was that the development impact of these projects was reduced, since the lending targets to key beneficiaries were not met. Among real sector projects, adverse business factors related to cyclicality, a downturn in the markets, or untested and flawed business models affected investment project performance. The third-most common negative factor was higher-than-expected competition, which affected the performance of 14 percent of projects. This led to investment projects missing operational targets and contributed to reduced operating margins and profitability. The fourth-most common adverse factor was the limited technical expertise and track record of sponsors and clients, affecting 13 percent of investment projects (see box 3.1 for examples by industry group).

Investment projects that accumulated several negative factors had lower development outcome ratings. The accumulation of several negative factors within one project created significant risks, which many projects were unable to overcome. The *RAP* team observed this for many investment projects in the Africa region, which had relatively low development outcome ratings. These investment projects were affected by adverse economic factors, high business risks, and low ability of sponsors and clients. These three factors were also the most common negative factors for investment projects in IDA and blend, and FCS countries.

Many private sector sponsors and clients reacted quickly to the changing economic landscape, showing remarkable resilience and adaptability during the pandemic. In the financial sector, most IFC clients contracted their loan portfolio and focused on asset quality issues. Many real sector project companies implemented cost-saving initiatives to increase efficiency and shore up margins. Others invested quickly in information technology solutions to facilitate remote work. Many companies rolled out online versions of their business lines, particularly companies in the higher education and food and consumer retail sectors. In a few cases, the pandemic increased the demand for clients' goods and services. For example, in the health care sector, investment project companies began manufacturing COVID-19 tests and vaccines, while project hospitals began treating COVID-19–affected patients.

The *RAP* team identified sponsor or client ability and technical expertise as a common factor that enabled proactive management to adapt quickly to the challenging environment. Strong sponsor and client ability and technical expertise contributed to better development outcomes. This factor positively influenced the performance of 30 percent of investment projects and was common for projects across all industry groups. Many sponsors can perform well in conducive operating environments, but strong and experienced sponsors can navigate challenging operating environments and identify mitigants to help projects survive. Indeed, strong sponsors were the decisive factor between investment projects on the borderline between mostly unsuccessful or mostly successful development outcome ratings. The main difference was that mostly successful investment projects relied on strong sponsors or clients to adapt to challenges.

There were other factors that supported investment project performance. Competitive business aspects supported the performance of 9 percent of investment projects, whereas favorable technology choices boosted the performance of 6 percent of projects. These two factors were most prevalent in real sector investment projects. Projects with clients that were market leaders or that increased their market share posted better operational and financial performance in 7 percent of the cohort. Strong financial capacity, capitalization, and leverage of sponsors aided the performance of 5 percent of projects. Collaboration and coordination among IFC investment and advisory teams, for example, by IFC providing technical assistance to sponsors,

enhanced the performance of 5 percent of investment projects, particularly in the financial sector.

Box 3.1. Examples of Supporting and Constraining Factors Affecting IFC Investment Project Performance, by Industry Group

Financial Institutions Group Projects

Supporting factors. Technical expertise and track record as a positive factor meant that the management of financial institutions was experienced and the financial institution had a historical strong performance in terms of earnings, asset quality, and risk management. Market share typically meant that the financial institution was the leader in the respective market (for example, banking, small and medium enterprise lending, microfinance, and housing), which gave it an edge over the competition. "Collaboration and coordination within the International Finance Corporation: advisory services and investment services" typically referred to joint International Finance Corporation investment services and advisory projects that helped improve the capacity of financial institutions, especially in the area of micro, small, and medium enterprise lending.

Constraining factors. Business factors meant that the financial institutions experienced declining performance of the targeted beneficiaries and moved away from lending to them (for example, micro, small, and medium enterprises; agribusiness; and affordable housing) as a result of the higher-risk profile. In a more adverse environment, the financial institutions turned toward making less risky loans to corporations or investing in government securities, which reduced development impact. Asset quality could mean that the financial institution did not sufficiently provision for bad loans or had deficiencies in credit risk management. Legal or regulatory factors meant that the regulatory environment became more adverse during project implementation, with interest rate caps or new policy requirements on financial institutions.

Infrastructure and Natural Resources Projects

Supporting factors. Legal or regulatory factors meant that the projects benefited from effective structuring of concession agreements and supportive government policies and initiatives. Technology meant that the projects benefited from technically and

(continued)

Box 3.1. Examples of Supporting and Constraining Factors Affecting IFC Investment Project Performance, by Industry Group (cont.)

commercially viable technology, with an edge over inefficient or costlier options. Pricing meant that the projects benefited from favorable tariffs or upward trends in market prices of their products.

Constraining factors. Business factors meant that the projects had flawed, untested, or fragile business models or experienced slowdown in market growth. Legal or regulatory factors meant that the projects were affected by failure to obtain the required licenses, an unexpected government decision to withhold value-added tax reimbursements, and disputes between the government and project company regarding the curtailment of fuel supply. Political factors meant the issues related to illiquid public sector offtaker with payment dependence on the government, inability of the government to meet its obligations in terms of fuel supply and offtake payments, delay in the commissioning attributed to the government, and regulatory changes because of the government's suspension of the privatization program.

Manufacturing, Agribusiness, and Services Projects

Supporting factors. Expansion meant that a project company benefited from expansion and market consolidation through acquisitions or organic growth, driving cost efficiencies and economies of scale. It could also mean that they had higher capital expenses or larger project scope than expected because of more investment. Relationship management meant that the project was a repeat deal with the same sponsors and gained from previous experience, or the International Finance Corporation had an active portfolio management and was flexible by helping the clients address their pressing needs in a more depressed market environment. Business factor as a favorable factor meant that the project gained from the increased market opportunity or its business model provided an edge over the competition.

Constraining factors. Environment and sustainability meant that the project had material shortcomings in meeting environmental and social requirements or that the client did not have in place some of the required important corporate policies. Business factors meant that the project company had shortcomings in the business model or suffered from unfavorable business and operating environment or industry cyclicality.

(continued)

Box 3.1. Examples of Supporting and Constraining Factors Affecting IFC Investment Project Performance, by Industry Group (cont.)

Disruptive Technologies and Funds Projects

Supporting factors. Technical expertise and track record as a positive factor meant fund managers with strong capacity or relevant experience. The environment and sustainability as a favorable factor meant the investment fund had high environmental and social and corporate governance standards.

Constraining factors. Project size meant that the fund was unable to reach its target size. This could be due to the fund manager's lack of experience or mean that the fund's investment thesis was too risky. Technical expertise and track record as an adverse factor typically meant that the fund manager lacked experience in private equity investing, in the specific fund target segment, or in emerging markets. Customers typically meant that the fund deviated from its investment strategy and invested in different types of portfolio companies than intended at approval. For example, the fund may have invested in developed countries rather than in emerging markets.

Source: Independent Evaluation Group.

Note: IFC = International Finance Corporation.

Performance factors could be divided into those within the control of IFC or its sponsors and those outside of their control. For example, an economic recession, the pandemic, or sudden changes in government regulations are out of the control of IFC and the project sponsors. However, sponsor ability, a client's market share, business factors, and IFC investment–advisory services collaboration can be within IFC and the project sponsors' control. Indeed, these controllable factors, which support project performance, featured more prominently in high-performing projects. Table 3.1 presents examples of potential measures that could be taken by IFC to mitigate adverse performance factors. For example, IFC or sponsors could mitigate adverse factors by conducting sound market analyses, enhancing the screening of sponsors, better assessing economic and business risks, providing advisory projects to strengthen sponsors' skills and capacity, and improving the delivery of additionality during implementation.

Table 3.1. Examples of Potential Mitigation Measures for IFC Investment Projects, by Industry Group

Adverse Performance Factor	Potential Mitigation Measures
Financial Institutions Group projects	
Business factors: Refers to a financial institution moving away from lending to the target beneficiaries (for example, micro, small, and medium enterprises; agribusiness; affordable housing; and so on) because of a more adverse environment, which will reduce development impact.	Mitigant: Provide technical assistance to the financial institution either before or during the implementation to enhance its capacity to increase or maintain lending to risky but highly developmental segments.
Integrity, transparency, fairness, and reputation: Refers to internal integrity due diligence issues with the sponsor affecting the project.	Mitigant: Conduct ongoing integrity due diligence to ensure that issues do not materialize. Proper supervision could help project teams react quickly to try to mitigate any adverse integrity due diligence issues during project implementation in a timely manner.
Loan factors: Refers to the entire facility not being drawn down or disbursed due to a change in the financial institution's strategy. It also refers to the loan tenor not being appropriate for the project or loan covenants not being appropriate or followed.	Mitigant: Carefully assess the financial institution's strategy and capacity at appraisal to ensure commitment to the project's development objectives (for example, micro, small, and medium enterprise lending). Ensure that the loan is properly priced or appropriate for the financial institution's purposes.
Infrastructure and Natural Resources projects	
Business factors: Refers to flawed, untested, or fragile business models or a slowdown in market growth.	Mitigant: Assess the viability of the business model during appraisal. Decline to invest when the business model is flawed. Provide additionality to assist the client in improving operations and practices.
Technical expertise and track record: Refers to the sponsors not measuring up to what the project was aiming to achieve without adequate operational and financial capacity, depth of management, and relevant experience.	Mitigant: Closely examine the sponsor's financial capacity, management depth, and relevant experience at appraisal. Provide additionality and active portfolio supervision if the sponsor decides to invest.

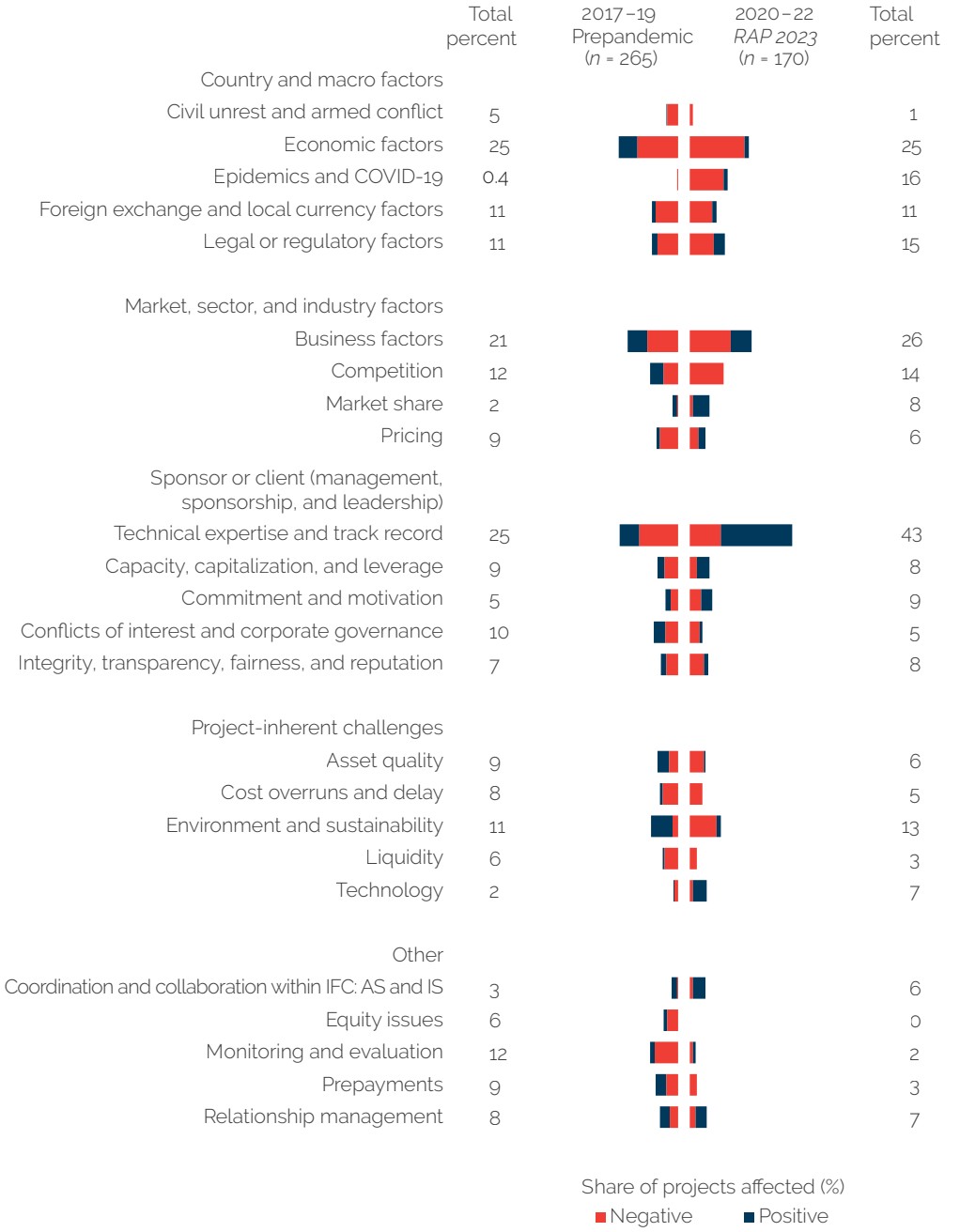
(continued)

Manufacturing, Agribusiness, and Services projects	
Environment and sustainability: Refers to the project having material shortcomings in meeting environmental and social requirements or the client not putting in place required corporate policies.	Mitigant: In case of corporate financing investments, ensure that the client has in place all required corporate policies and that all its businesses comply with environmental and social performance standards. The cost of environmental and social improvements needs to be estimated at appraisal and included in the project cost if needed. Provide additionality to improve the client's environmental and social practices.
Business factors: Refers to shortcomings in the business model or unfavorable business and operating environment or industry cyclicality.	Mitigant: Assess the viability of the business model during appraisal. Provide additionality to clients in improving their operations and practices.
Disruptive Technologies and Funds projects	
Project size: Refers to a fund not reaching its target size, potentially due to fund manager's lack of experience or riskiness of investment thesis.	Mitigant: Provide additionality by assisting fund manager in fundraising. Conversely, decline to invest in the fund if the fund is unable to reach the minimum capital.
Technical expertise and track record: Typically refers to a fund manager lacking experience in private equity investing, in the specific fund target segment, or in emerging markets.	Mitigant: Provide additionality through technical assistance to both the fund manager and the downstream portfolio companies to help make the fund successful.
Customers: Typically refers to fund deviation from its investment strategy and investing in different types of portfolio companies than intended at approval.	Mitigant: Through position on the advisory committee, voice objections to any unnecessary deviations in the fund strategy or decline to participate in investments that are not in line with the investment thesis as presented at approval.

Source: Independent Evaluation Group.
Note: IFC = International Finance Corporation.

Investment projects from the CY20–22 cohort operated in more challenging country and market conditions than did prepandemic projects. This *RAP* compared performance factors of CY20–22 investment projects with those of CY17–19 projects (see appendix A and the report by Bravo et al. 2023 for the methodological details of applying supervised machine learning to the analysis of CY17–19 projects). Except for the pandemic and related effects, the key factors affecting CY20–22 investment project performance were broadly the same as those affecting CY17–19 project performance. That said, the investment projects in the *RAP* cohort were more negatively affected by (i) adverse economic factors; (ii) high business risks; (iii) unforeseen epidemics, including COVID-19; and (iv) higher-than-expected competition. The environment and sustainability factor was also more prevalent in CY20–22 investment projects. In contrast, M&E issues and unfavorable market pricing were more prevalent negative factors for the prepandemic investment projects (figure 3.5). However, strong client or sponsor ability, technical expertise, and experience aided the performance of investment projects in the *RAP* cohort. This factor was also prevalent in high-performing investment projects from CY20–22, demonstrating that strong clients and sponsors were able to effectively cope with challenges posed by the pandemic.

Figure 3.5. Factors Affecting IFC Investment Project Performance: The Prepandemic Cohort Compared with the *RAP 2023* Cohort



Source: Independent Evaluation Group.

Note: The factor identification for calendar years 2020–22 projects was based on human thinking, whereas for calendar years 2019–21 prepandemic projects, it was based on machine learning. Positive = the identified factor aided the project performance. Negative = the identified factor constrained the project performance. AS = advisory services; IFC = International Finance Corporation; IS = investment services; *RAP* = *Results and Performance of the World Bank Group*.

Outcome Types of International Finance Corporation Investment Projects

This *RAP* built on *RAP 2021* outcome type analysis with a deep-dive analysis and found that IFC investment project objectives were highly outcome oriented. The deep dive examined the intended outcomes from 170 IFC investment projects in the *RAP* cohort. The deep dive identified 848 outcomes across 13 different outcome types (see appendix A for the outcome typology). IFC's investment project outcomes fall into two broad categories: project-level outcomes and market-level outcomes. Project-level outcomes are those with direct and indirect effects on stakeholders, the economy, and the environment. Market-level outcomes are derived effects, or those that catalyze systemic changes beyond the project's effects.[11] Projects reviewed in the deep dive pursued an average of 5 different outcomes, consisting of 4 project-level outcomes and 1 market-level outcome (see box 3.2 for examples of project- and market-level outcomes by industry group). Overall, all reviewed IFC investment projects pursued project-level outcomes, and 74 percent pursued market-level outcomes, confirming the *RAP*'s hypothesis that IFC investment projects were focused on higher-level outcomes such as market-level outcomes. Every IFC project pursued the project-level outcome type of improved access to goods and services. Other prevalent project-level outcomes were increased employment and quality and affordability of goods and services. The most common market-level outcome—competition in the market—was prevalent in 58 percent of projects, whereas sustainability in the market (which refers to clients' adoption of climate-friendly and environmentally and socially sustainable products, practices, and technologies) was the second most prevalent market-level outcome.

Box 3.2. Examples of IFC Investment Project-Level and Market-Level Outcomes, by Industry Group

Financial Institutions Group

Project-level outcome: Increase in outstanding small and medium enterprise loans, increase in share of microfinance loans, and reduction in nonperforming loans ratio.

(continued)

Box 3.2. Examples of IFC Investment Project-Level and Market-Level Outcomes, by Industry Group (cont.)

Market-level outcome: Demonstration of the viability of lending to microborrowers or small and medium enterprises, deepening of financial markets, and fostering increased competition in the banking sector.

Infrastructure and Natural Resources

Project-level outcome: Increase in renewable energy generation, improvement in information technology infrastructure, increase in access and use of mobile telecommunication services, and number of passengers with access to the road.

Market-level outcome: Diversification of energy mix and increased competition in the information and communication technology sector.

Manufacturing, Agribusiness, and Services

Project-level outcome: Increase in affordable housing supply, increase in purchases from domestic suppliers, increase in quality or affordability of health care services, and increase in tax payments.

Market-level outcome: Demonstration effect on the local agribusiness industry; demonstration of viability of green buildings and promotion of replication; enhanced environmental, social, and governance standards to serve as a corporate role model.

Disruptive Technologies and Funds

Project-level outcome: Percentage of fund investee companies with growth in revenue and returns, increase in job creation at investee companies, and increase in access to information and communication technology services.

Market-level outcome: Demonstration effect through raising of follow-on fund and facilitation of investee companies' emergence as regional players.

Source: Independent Evaluation Group.

Note: IFC = International Finance Corporation.

Outcome achievement rates were relatively low. IFC fully achieved 45 percent of its 693 stated project-level outcomes and partially achieved 22 percent. The highest achievement rate among common project-level outcomes was improved sales and profitability of enterprises, which was fully achieved 67 percent of the time. Other common project-level outcomes with high full achievement rates were greenhouse gas reductions at 63 percent, enhanced capacity of final beneficiaries at 57 percent, and enhanced environmental and social standards of the client at 52 percent. IFC fully achieved 43 percent of its 155 stated market-level outcomes and partially achieved 21 percent. Among these, resilience in the market had the highest full achievement rate at 47 percent, whereas competition in the market, which was the most common market-level outcome in projects, had a full achievement rate of 45 percent (see table 3.2 for achievement rates by outcome type).

Table 3.2. IFC Outcome-Type Performance: Achievement Rate

Outcome Type	Outcomes (no.)	Outcome Achieved (fully; %)	Outcome Achieved (partially; %)	Outcome Achieved (fully + partially; %)
Project-level outcomes				
1.1 - Access to goods and services	242	44	30	74
1.1.1 - Access to goods and services (MSME)	50	46	22	68
1.1.2 - Access to goods and services (female)	13	62	15	77
1.1.3 - Access to goods and services (customers)	88	43	32	75
1.1.4 - Access to goods and services (miscellaneous)	66	39	33	72
1.1.5 - Access to goods and services (direct client level)	25	44	36	80

(continued)

Outcome Type	Outcomes (no.)	Outcome Achieved (fully; %)	Outcome Achieved (partially; %)	Outcome Achieved (fully + partially; %)
1.2 - Quality and affordability of goods and services	104	47	14	61
1.2.1 - Quality of goods and services	37	46	14	60
1.2.2 - Affordability of goods and services	53	47	13	60
1.2.3 - Improved productivity and efficiency of the direct client	14	50	21	71
1.3 - Enhanced capacity of final beneficiaries	23	57	13	70
1.4 - Improved living standards (earnings) of individuals	5	0	0	0
1.5 - Improved sales and profitability of enterprises	15	67	27	94
2.1 – Suppliers and distributors reached	12	50	25	75
2.2 - Improved capacity of suppliers and distributors	5	40	20	60
2.3 - Improved sales and profitability of suppliers and distributors	28	39	32	71
3.1 - Increased employment	94	40	23	63
3.2 - Improved capacity and skills	11	45	27	72
3.3 - Improved earning of employees	0	n.a.	n.a.	n.a.

(continued)

Outcome Type	Outcomes (no.)	Outcome Achieved (fully; %)	Outcome Achieved (partially; %)	Outcome Achieved (fully + partially; %)
4.1 - Increased transfers to the government	32	31	28	59
5.1 - Increased money spent and transfer to the communities	5	60	0	60
6.1 - Enhanced E&S standards of the client	54	52	17	69
6.2 - Greenhouse gas reduction	32	63	6	69
6.3 - Efficient use of resources	11	55	9	64
7.1 - Gross value added	5	40	60	100
7.2 - Induced or indirect employment	8	25	0	25
7.3 - Export sales	4	50	0	50
8.1 - Governance	3	67	0	67
Total project-level outcomes	693	45	23	68
Market-level outcomes				
9 - Competition in the market	98	45	20	65
10 - Resilience in the market	17	47	29	76
11 - Integration in the market	12	33	25	58
12 - Inclusiveness in the market	8	38	13	50
13 - Sustainability in the market	20	40	20	60

(continued)

Outcome Type	Outcomes (no.)	Outcome Achieved (fully; %)	Outcome Achieved (partially; %)	Outcome Achieved (fully + partially; %)
Total market-level outcomes	155	43	21	65
Total outcomes	848	45	22	67

Source: Independent Evaluation Group.

Note: Of project-level outcomes, 8 percent were considered not achieved because the results could not be verified. Of market-level outcomes, 7 percent were considered not achieved because their results could not be verified. E&S = environmental and social; IFC = International Finance Corporation; MSME = micro, small, and medium enterprise; n.a. = not applicable.

The achievement of market-level outcomes was almost as high as the achievement of project-level outcomes. These results did not fully confirm this *RAP*'s hypothesis that market-level outcomes were more difficult to achieve than project-level outcomes. This hypothesis was supported by findings from *RAP 2021*, which found that market-level outcomes are more difficult to achieve "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" (World Bank 2021, 51). However, this *RAP* shows (see table 3.2) that reviewed investment projects fully achieved project-level outcomes only 2 percent more often than they achieved market-level outcomes (45 percent compared with 43 percent).

Monitoring data were not available for a significant number of total outcomes. This confirms the *RAP*'s hypothesis that IFC's result measurement indicators are not fully adequate to measure outcome achievement. This is consistent with *RAP 2021*, which states that "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" (World Bank 2021, xv). Most of the investment projects in the *RAP 2023* cohort were not subject to an AIMM assessment at their approval and continued to be monitored in the Development Outcome Tracking System. In many cases, IFC or IEG used other information sources, where available, to validate project outcome claims.

Some outcomes could not be verified because of a lack of appropriate results measurement indicators and evidence, which depressed outcome achievement rates. Eight percent of total outcomes could not be verified because of a

lack of evidence and were coded as "cannot be verified," including 8 percent of project-level outcomes and 7 percent of market-level outcomes. Of the 65 outcomes that could not be verified, 91 percent were not tracked by IFC in any monitoring system. Some of the most common reasons an outcome could not be verified were as follows: (i) the project did not have an indicator to track the outcome, (ii) the client did not report relevant information, (iii) there was insufficient evidence to measure achievement, (iv) there was no clarity in how to measure the outcome, (v) the result could not be attributed to the project, or (vi) it was too early to tell.

Relationship between Outcomes and Project Performance Ratings

Investment project development outcome ratings were related to the achievement rates of project- and market-level outcomes. According to the XPSR guidelines, development outcome ratings of IFC investment projects are assigned at the project level and subdimension level but not at the project outcome level. Therefore, this *RAP* expanded the *RAP 2021* outcome type analysis by comparing outcome achievement to individual project development outcome ratings. This analysis showed that IFC investment projects that achieved more of their outcomes also had higher development outcome ratings. This relationship was particularly strong for market-level outcomes (table 3.3). For the *RAP 2023* cohort of 170 investment projects, development outcome ratings decreased, along with lower outcome achievement rates, for both project- and market-level outcomes. Highly successful projects achieved 100 percent of their project- and market-level outcomes. Development outcome ratings declined in tandem with lower outcome achievement. Highly unsuccessful projects achieved only 8 percent of their outcomes. This indicates a clear link between outcome achievement levels, especially for market-level outcomes, and development outcome rating (see appendix F for details on outcome type analysis). Project business performance (PBP) and environmental and social effects are components of development outcome; therefore, lower PBP and environmental and social performance were also associated with lower investment project development outcome ratings.

Table 3.3. IFC Investment Project Development Outcome Ratings and Underlying Outcome Achievement Rates

Development Outcome Rating	Total Projects (no.)	Total Outcomes (no.)	Overall Weighted Achievement Rate (%)	Project-Level Outcome Weighted Achievement Rate (%)	Market-Level Outcome Weighted Achievement Rate (%)	Project Business Performance Average Rating	Environmental and Social Effects Average Rating
Highly successful	3	18	100	100	100	4.0	3.0
Successful	35	161	87	86	90	3.4	3.0
Mostly successful	46	250	73	75	68	2.8	2.8
Mostly unsuccessful	39	198	49	50	45	2.2	2.5
Unsuccessful	32	139	23	25	9	1.4	2.5
Highly unsuccessful	15	82	8	8	8	1.1	1.8

Source: Independent Evaluation Group.

Note: Outcome achievements in projects are measured with the following weights: outcome achieved = 1, partially achieved = 0.5, not achieved = 0, and cannot verify = 0. Project business performance and environmental and social effects ratings' numerical values are as follows: excellent = 4, satisfactory = 3, partly unsatisfactory = 2, and unsatisfactory = 1. IFC = International Finance Corporation.

An investment project's level of outcome achievement was the main difference in influencing the development outcome rating for borderline projects. As table 3.3 shows, the weighted outcome achievement rate of investment projects rated mostly successful was 73 percent compared with only 49 percent for investment projects rated mostly unsuccessful. Investment projects rated mostly successful achieved 68 percent of market-level outcomes compared with only 45 percent for investment projects rated mostly unsuccessful. Financial performance was also an important factor for borderline projects as the difference in their PBP rating was one full rating difference (satisfactory versus partly unsatisfactory). Investment projects rated mostly successful had an average PBP rating of 2.8 (closer to satisfactory than partly unsatisfactory), whereas investment projects rated mostly unsuccessful had an average PBP rating of 2.2 (closer to partly unsatisfactory than satisfactory).

[1] The *Results and Performance of the World Bank Group* cohort included 221 International Finance Corporation (IFC) investment projects in calendar years 2020–22 Expanded Project Supervision Report (XPSR) programs evaluated and validated as of June 30, 2023. These projects were selected in the random representative sample from the net approvals population of projects approved during calendar years 2012–17.

[2] The *Results and Performance of the World Bank Group* cohort was composed of 175 IFC advisory projects in fiscal years 2020–22 Project Completion Report programs, evaluated and validated as of June 30, 2023. These projects were selected in the random representative sample from the population of self-evaluated projects, which were approved during fiscal years 2009–20. Fifteen of these projects were flagged as upstream projects.

[3] All investment projects in calendar years 2020–22 XPSR programs were evaluated and validated after March 2020.

[4] The COVID-19 impact data were provided by IFC at the time of sampling. For each project in the raw population for sampling, IFC provided to the Independent Evaluation Group the detailed information regarding the project's operational status and the magnitude of COVID-19 impact (low, medium, high) based on IFC's own assessment at the time. All projects deemed as highly affected by COVID-19 by IFC were automatically deferred for evaluation, while those with medium or low COVID-19 impact were reviewed on a project-by-project basis to decide their readiness for evaluation.

[5] Projects that were deferred due to COVID-19 represented 15 percent of the raw population for calendar years 2020–22 XPSR programs.

[6] The ratings analysis for all performance indicators, including for work quality of IFC, is based on the Independent Evaluation Group's independent ratings, unless stated otherwise. While IFC conducts a self-assessment of IFC work quality, there is no requirement for IFC to self-rate IFC work quality in the XPSRs.

[7] The Pearson correlation coefficient between development outcome and IFC's overall work quality and the front-end work quality was 0.62 and significant at 95 percent confidence level. Meanwhile, the correlation coefficient of the development outcome and IFC's supervision work quality was 0.33.

[8] The cited gaps between anticipated and realized additionalities considered both fully and partially realized additionalities.

[9] The Pearson correlation coefficient between development outcome and IFC additionality was 0.64 and significant at 95 percent confidence level.

[10] Among investment projects with low loan outcomes, 42 percent had prepayments.

[11] The IFC 3.0 strategy explicitly prioritizes creating markets, which falls into the market-level category (IFC 2016, 2019).

4 | Multilateral Investment Guarantee Agency Results and Performance

Multilateral Investment Guarantee Agency (MIGA) guarantee projects in the *Results and Performance of the World Bank Group* cohort were exposed to, and affected by, the COVID-19 pandemic. On average, these projects were exposed to the pandemic conditions for 27 percent of their active project lives. COVID-19–related lockdowns and economic slowdowns reduced the demand for services of MIGA's public transportation projects.

MIGA's overall development outcome success ratings remained stable in fiscal years 2017–22, with 72 percent of guarantee projects rated satisfactory or better. However, these ratings were slightly lower over the last three years, partially reflecting the more challenging operating environment.

The reviewed MIGA guarantee projects did not have formal procedures for adjusting guarantee projects' development objectives and outcome targets in response to crises such as COVID-19.

MIGA's role and contribution ratings were high, with the most significant contributions in environmental and social areas and some project risk mitigation.

Half of guarantee projects in fiscal years 2017–22 had work quality shortcomings, particularly the underwriting and structuring work quality.

Local currency depreciations and cost overruns and construction delays were the two most common factors that undermined project performance.

Some MIGA guarantee projects adapted to the pandemic, showing resilience and flexibility. Capable sponsors and favorable legal and regulatory aspects helped effectively counter adverse factors.

This *Results and Performance of the World Bank Group*'s deep-dive analysis shows that MIGA guarantee project objectives are highly outcome oriented. However, guarantee projects achieved the outcomes defined at approval at a low rate. MIGA guarantee projects that achieved more of their intended outcomes had higher development outcome ratings.

The monitoring and evaluation of foreign investment–level outcomes is challenging; therefore, the Independent Evaluation Group and MIGA could not measure or validate some project development outcomes.

This chapter presents trends and patterns of MIGA guarantee project performance. It also describes MIGA guarantee projects' operating environment and how potential market, or efficiency, tests can contribute to project success. The chapter also explores the development outcomes that underpin MIGA's outcome ratings and examines key factors associated with project implementation and performance. It should be noted that only 60 projects were evaluated and validated in FY17–22 as of June 30, 2023;[1] therefore, individual project ratings can make a large difference in the portfolio's overall success rates. As such, this chapter's results should be interpreted with caution.

Project Exposure to the COVID-19 Pandemic and Sample Selection Bias

The implementation of half of MIGA guarantee projects in the last three years was affected by COVID-19 to at least some extent. On average, these MIGA guarantee projects were exposed to the pandemic for 27 percent of their active project lives. In addition, there were delays in the delivery of some MIGA self-evaluations, which limited the number of guarantee projects analyzed in the *RAP* cohort.[2] Delays in delivery of self-evaluations were related to client reporting delays and challenges in visiting project sites, which constrained information gathering and required more time to fill information gaps. The relatively small sample size and the fact that more self-evaluations of projects that took place during the pandemic will be completed at a later date indicate that this *RAP*'s analysis provides only limited and preliminary insights on the pandemic's effects on MIGA guarantee projects.

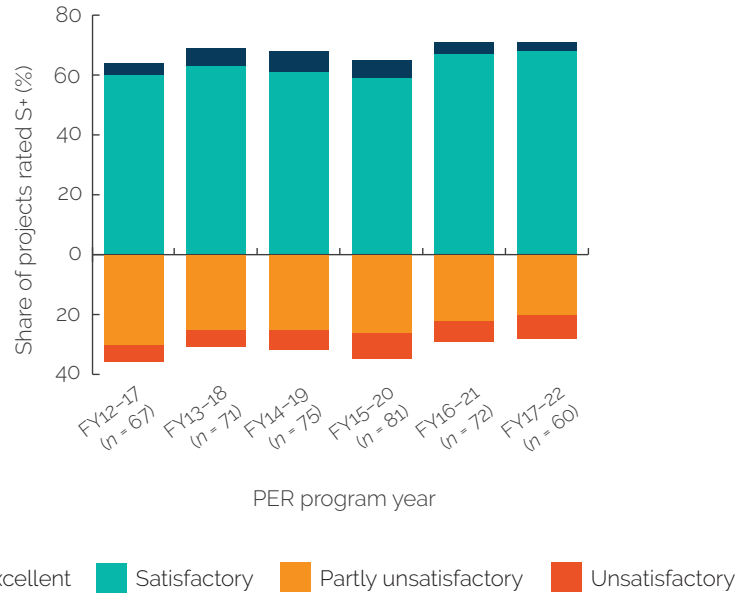
COVID-19 exposed MIGA guarantee projects to unforeseen implementation challenges. As discussed elsewhere, the pandemic caused lockdowns, economic slowdowns, and supply chain disruptions. Lockdowns, in particular, reduced consumer needs for public transportation and fuel products, thereby diminishing the demand for some of MIGA's Infrastructure guarantee projects. Some hospital projects in the Agribusiness and General Services sector also experienced reduced demand for elective health care services, which were considered nonessential during the pandemic.

As with IFC investment projects, MIGA also has no formal procedures for changing its guarantee project–related objectives or targets when confronted with crises. MIGA-guaranteed private sector projects must be efficient and competitive to be commercially viable while simultaneously delivering development outcomes. In other words, MIGA guarantee projects must deliver the double bottom line of achieving financial sustainability and development impact. Moreover, all MIGA guarantee projects are required to comply with MIGA's environmental and social performance standards, thereby achieving a triple bottom line. MIGA guarantee projects can adapt to changing market conditions, such as those caused by the COVID-19 pandemic. However, the existing framework does not include the adjustment of such originally envisioned development objectives and targets, considering ex post changed market conditions, such as those caused by the COVID-19 pandemic. According to MIGA, the current Impact Measurement and Project Assessment Comparison Tool (IMPACT) framework allows an adjustment of the expected impact claims when processing contract modifications. However, IEG has not yet evaluated any projects with IMPACT-tracked development objectives.

Project Performance Rating Trends

Development outcome success ratings of MIGA guarantee projects remained stable. On a six-year rolling basis, 72 percent of MIGA guarantee projects in FY17–22 were rated satisfactory or better, the same level as in FY16–21 (figure 4.1). The composition of granular project development outcome ratings in FY17–22 remained largely the same as in FY16–21. That said, these development outcome success rates were slightly lower over the last three years because of a more challenging operating environment. A larger share of negatively rated projects in FY20–22 offset higher-performing projects in FY17–19 (see appendix H, figure H.2).

Figure 4.1. MIGA Guarantee Project Development Outcome Ratings



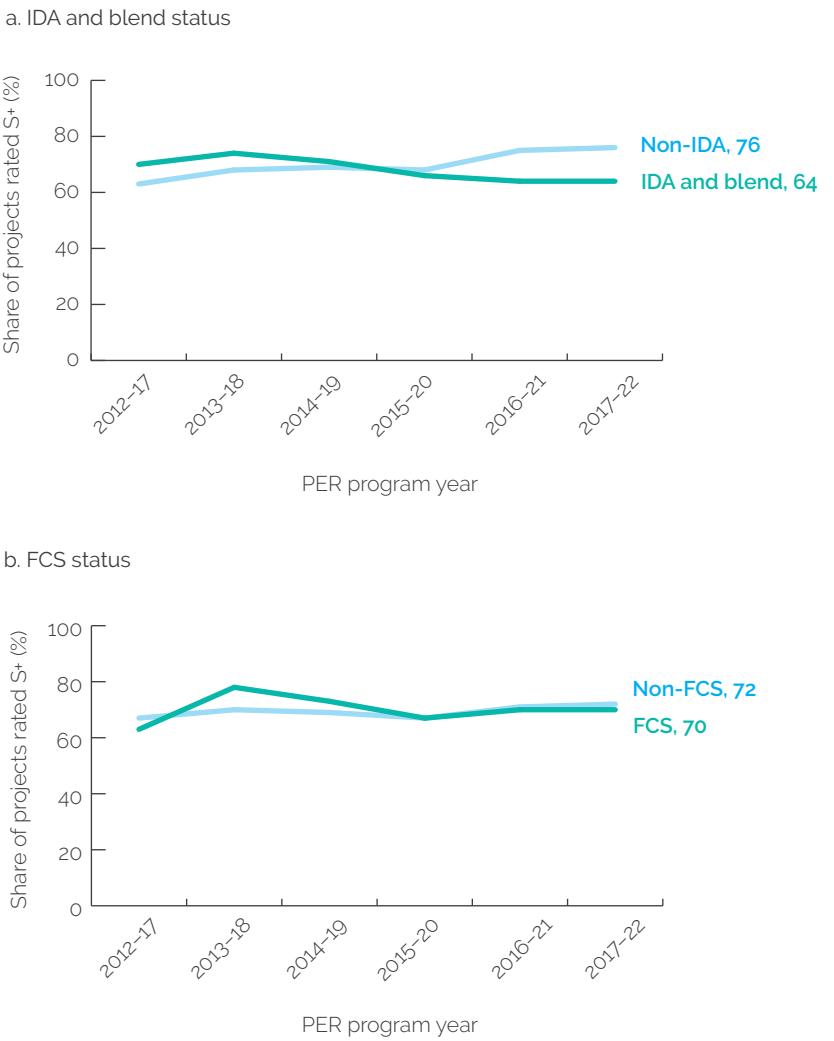
Source: Independent Evaluation Group.

Note: The Multilateral Investment Guarantee Agency Project Evaluation Report guidelines were changed in FY19, replacing a four-point scale for development outcome ratings with a six-point one. The six-point rating scale, applied to projects starting in FY20, was converted to a four-point one as follows: highly successful = excellent; successful and mostly successful = satisfactory; mostly unsuccessful = partly unsatisfactory; and highly unsuccessful and unsuccessful = unsatisfactory. FY = fiscal year; MIGA = Multilateral Investment Guarantee Agency; PER = Project Evaluation Report; S+ = satisfactory or better.

The performance gap between MIGA guarantee projects in IDA and blend countries and those in non-IDA countries largely stayed the same during FY17–22. MIGA's overall development outcome ratings of guarantee projects in both IDA and blend and non-IDA countries remained stable (figure 4.2, panel a). However, the performance of IDA and blend projects, representing 37 percent of MIGA's evaluated portfolio, continued to be lower than that of non-IDA projects, with 64 percent rated satisfactory or better for development outcome in FY17–22. In contrast, guarantee projects in non-IDA countries, representing 63 percent of MIGA's evaluated portfolio, maintained satisfactory or better ratings for 76 percent of projects in FY17–22. This is important because MIGA's FY21–23 strategy priority is to deepen its commitment to IDA and blend countries. Indeed, the share of MIGA's overall portfolio operating in these countries has significantly increased from 24 percent in FY19 to 65 percent in FY22. The evaluated portfolio of MIGA guarantee projects in FCS

countries was generally small, with 10 or fewer projects per period. That said, the percentage of those rated satisfactory or better for overall development outcome remained stable at 70 percent between FY16–21 and FY17–22 (figure 4.2, panel b). This performance in FCS countries was on par with performance in non-FCS projects.

Figure 4.2. MIGA Guarantee Project Development Outcome Success Ratings, by IDA and Blend and FCS Status



Source: Independent Evaluation Group.

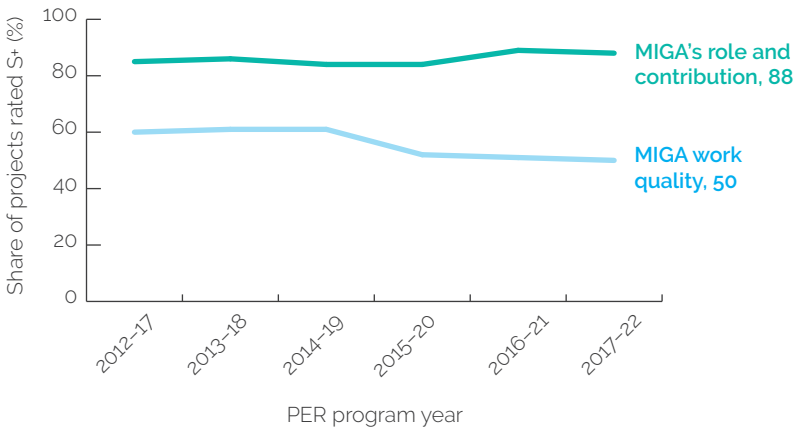
Note: FCS = fragile and conflict-affected situation; IDA = International Development Association; MIGA = Multilateral Investment Guarantee Agency; PER = Project Evaluation Report; S+ = satisfactory or better.

As with IFC investment projects, MIGA guarantee projects have many market-related hurdles to overcome to be successful. As an insurer, MIGA

guarantees a sponsor's investment, but it is the sponsor, not MIGA, that deals directly with the project enterprise. Therefore, MIGA is one step removed from the project, limiting its influence over project design, structuring, and implementation. By contrast, MIGA stays closely involved in a project's environmental and social areas and its risk reduction. In IDA and FCS countries, MIGA is constrained by the limited capacity and experience of available sponsors. Moreover, the fact that project sponsors seek MIGA's political risk guarantees suggests that political risk is another challenge in these countries. These risks make sponsors reluctant to invest without MIGA guarantees.

MIGA achieved high success rates in carrying out its expected role and contribution. The share of guarantee projects with satisfactory ratings for MIGA role and contribution was 88 percent in FY17–22, the same level as in FY16–21 (figure 4.3). In fact, MIGA's role and contribution ratings were generally high across the entire portfolio. For example, MIGA achieved its expected role and contribution in 80 percent of guarantee projects in FCS countries. Moreover, MIGA's role and contribution ratings in IDA and blend countries were on par with those same ratings in non-IDA countries. MIGA's role and contribution was most significant in environmental and social areas and risk reduction.

Figure 4.3. MIGA Work Quality and Role and Contribution Success Ratings



Source: Independent Evaluation Group.

Note: MIGA = Multilateral Investment Guarantee Agency; PER = Project Evaluation Report; S+ = satisfactory or better.

Half of MIGA guarantee projects in FY17–22 continued to have work quality shortcomings. Sixty percent of MIGA guarantee projects for the six-year rolling average over FY12–17 were rated satisfactory or better, but the share fell to 51 percent in FY16–21 and to 50 percent in FY17–22 (see figure 4.3). The decline was even more evident when looking at three-year rolling averages, falling from 56 percent rated satisfactory or better in FY15–17 to 48 percent in FY19–21 and just 43 percent in FY20–22. MIGA work quality rating was correlated with the development outcome rating in 75 percent of guarantee projects in FY17–22. Twenty-seven percent of these cases had low development outcomes, which were associated with MIGA's weak work quality.

Factors Affecting Project Implementation and Performance

Several factors besides COVID-19 negatively affected the performance of recent MIGA guarantee projects. As with IFC, this *RAP* conducted a deep-dive analysis of 16 MIGA guarantee projects validated from FY20 to FY22 and identified the top factors affecting the guarantee project performance.[3] These factors could have either a negative or positive influence on project performance (see appendix A for the definitions of different factors). Ten of the 16 projects had high development outcome ratings, and 6 projects had low development outcome ratings. Cost overruns and construction delays and unfavorable foreign exchange rates were the two most common adverse factors, with each affecting 25 percent of the analyzed projects. Cost overruns and construction delays undermined the financial and economic performance of 4 projects in the energy, public transportation, and extractive industry sectors. Foreign exchange factors undermined the financial results of 4 MIGA guarantee projects in the health sector. Another factor—increased competition—negatively affected 3 projects, or 19 percent, in the energy and telecom sectors by weakening their operational and financial results. Inadequate M&E prevented 3 Infrastructure projects from verifying some of their intended development impacts. MIGA guarantee projects were also challenged by other factors to a lesser extent, including adverse economic issues, increased political risk, inadequate market assessments, and the low technical expertise and track record of sponsors and project enterprise management.

Some MIGA guarantee projects adapted to the pandemic. For example, MIGA hospital projects in the Agribusiness and General Services sector adapted to the pandemic by assisting the government's response to the pandemic's emerging medical demands. The projects offered COVID-19 treatment, which increased demand for these projects' operations.

Several factors positively influenced MIGA guarantee project success, particularly strong sponsor technical expertise and favorable legal and regulatory factors. These two factors influenced 31 percent of projects in the MIGA *RAP* cohort, especially the four health sector projects. These greenfield hospital public-private partnership (PPP) projects benefited from having sponsors with financial capacity, competent management, and relevant experience in implementing PPP projects. The payment mechanism in these PPP projects protected sponsors from the downside risk of a local currency depreciation. It was also the strong technical expertise and track record of sponsors and project enterprise management along with favorable legal and regulatory factors that separated mostly successful projects from mostly unsuccessful projects. These two mitigating factors enabled these mostly successful projects to cope with the pandemic and other crises and achieve largely positive results. Other factors aiding MIGA's project performance in FY20–22, but to a lesser extent, were as follows: (i) a high market share for the project enterprise's business (13 percent), (ii) favorable business aspects (13 percent), (iii) positive environmental and social aspects (13 percent), and (iv) savings in project costs and construction times (13 percent). See box 4.1 for examples of how these factors affected project performance across sectors.

Box 4.1. Examples of Supporting and Constraining Factors Affecting MIGA Guarantee Project Performance, by Sector

Finance and Capital Markets Projects

Supporting factors. Positive technical expertise and track record meant that the financial institution management was strong and the financial institution had a strong performance in terms of total loan growth, earnings, capital adequacy, and liquidity. Market share meant that the financial institution was the leader in the small and medium enterprise lending market, which gave it an edge over the competition.

Constraining factors. Weak earnings and profitability meant that the financial institution had weaker-than-expected loan growth and financial performance, including deteriorated asset quality, lower profitability, and tighter liquidity. Inadequate market assessment meant that MIGA did not adequately assess the financial institution's commitment and capacity to achieve the project's intended objectives, given that the ultimate beneficiaries of the MIGA-supported financing were not realized as expected, thus reducing the development impact.

Infrastructure and Energy and Extractive Industries Projects

Supporting factors. Positive project cost and construction times meant that the project construction was under budget and that implementation was on time.

Constraining factors. Adverse competition meant that the project enterprise experienced a highly competitive market that depressed revenues, suffered from overcapacity in the sector, or faced competition from more efficient new entrants in the market. Cost overruns and construction delays meant that the project suffered from implementation delays and cost overruns, which negatively affected the project's financial and economic returns. Monitoring and evaluation issues meant that the project had shortcomings in monitoring and evaluation, such as the lack of quantified baseline or targets and information on actual results, preventing the verification of achievement of its intended development impacts.

(continued)

Box 4.1. Examples of Supporting and Constraining Factors Affecting Performance of MIGA Guarantee Projects, by Sector (cont.)

Agribusiness and General Services Projects

Supporting factors. The high technical expertise and track record meant that the sponsors had financial capacity, relevant experience in implementing public-private partnership projects, and competent management that ensured high-quality operations and maintenance. The positive legal and regulatory factors meant that the payment mechanism in public-private partnership projects protected the sponsors from the downside risk of a depreciation of the local currency.

Constraining factors. The adverse foreign exchange and local currency factors meant that depreciation of the local currency and resulting foreign exchange losses negatively affected financial results of the project enterprises.

Source: Independent Evaluation Group.

Note: MIGA = Multilateral Investment Guarantee Agency.

Similar to IFC, the deep-dive analysis found that performance factors affecting MIGA guarantee projects could be divided into those within the control of MIGA, its sponsor, or the project enterprise, and those outside of their control. For example, economic issues, the COVID-19 pandemic, and legal and regulatory changes were not foreseeable and could be considered outside of the control of MIGA, the sponsor, or the project enterprise. However, other factors could be considered within their control and thus could have been mitigated. Among these factors, the technical expertise of sponsors, market assessment, and M&E could be considered within their control. See table 4.1 for examples of potential measures that could mitigate adverse performance factors.

Table 4.1. Examples of Potential Mitigation Measures for MIGA Guarantee Projects, by Sector

Adverse Performance Factor	Potential Mitigation Measures
Finance and Capital Markets Projects	
Inadequate market assessment: Refers to the inadequate assessment of a financial institution's commitment and capacity to achieve the project's intended objectives. In such cases, the beneficiaries of the financing are not supported as expected, reducing development impact.	Mitigant: Better define development impact objectives at appraisal by clarifying the purpose and use of proceeds of guaranteed facilities and establishing appropriate development impact indicators.
Infrastructure and Energy and Extractive Industries Projects	
Monitoring and evaluation: Refers to the project having shortcomings in monitoring and evaluation, such as the lack of quantified baseline or targets and information on actual results, preventing the verification of achievement of its intended development impacts.	Mitigant: Establish appropriate mechanisms for development impact data gathering in guarantee projects, where the project enterprise is not a direct signatory to supported financing agreements.
Agribusiness and General Services Projects	
Foreign exchange and local currency: Refers to depreciation of the local currency that results in foreign exchange losses that negatively affect the financial results of the project enterprise.	Mitigant: Identify and assess the potential impact of foreseeable macroeconomic developments, including depreciation of the local currency, that may increase the size of the government's financial obligations and assess whether the government will be willing and have the capacity to pay the increased obligations.

Source: Independent Evaluation Group.

Note: MIGA = Multilateral Investment Guarantee Agency.

Outcome Types of Multilateral Investment Guarantee Agency Guarantee Projects

This *RAP* continued the *RAP 2021* outcome type analysis and found that MIGA project objectives were highly outcome oriented. The deep-dive analysis looked at 78 outcomes within 13 outcome types, in 16 MIGA guarantee projects validated from FY20 to FY22 (see appendix A for the

outcome typology). MIGA guarantee project outcomes fall into two broad categories: project-level outcomes and foreign investment–level outcomes. Of the 78 outcomes, 62 were intended project-level outcomes and 16 were intended foreign investment–level outcomes. Reviewed projects established, on average, 5 outcome types per project—this usually included 4 project-level outcomes and 1 foreign investment–level outcome. Overall, all projects pursued project-level outcomes, and 81 percent pursued at least 1 foreign investment–level outcome. This confirms the *RAP 2023* Concept Note's hypothesis that MIGA guarantee projects were outcome oriented or were focused on higher-level outcomes, such as foreign investment–level outcomes beyond the project level. These outcome types reflected MIGA's focus on larger infrastructure projects for the health care, energy, telecommunications, and transportation sectors in FY20–22. The most prevalent foreign investment–level outcome was market development, found in 56 percent of projects. For project-level outcomes, all MIGA guarantee projects included improved access to goods and services. Other prevalent project-level outcomes included increased employment and quality and affordability of goods and services, which were present in 63 percent and 50 percent of projects, respectively.

MIGA's outcome achievement rates were relatively low. Of the 78 outcomes defined at approval, MIGA guarantee projects fully achieved 50 percent and partially achieved 22 percent. Of the 62 project-level outcomes defined at approval, projects fully achieved 55 percent and partially achieved 21 percent (table 4.2). Of the 16 foreign investment–level outcomes, projects fully achieved 31 percent and partially achieved 25 percent. This shows that MIGA guarantee projects had a higher probability of achieving project-level outcomes than foreign investment–level outcomes, thereby confirming the *RAP 2023* hypothesis and *RAP 2021* findings that foreign investment–level outcomes are more challenging to achieve.

Table 4.2. MIGA Outcome Type Performance: Achievement Rate

Outcome Type	Outcomes (no.)	Outcome Achieved (fully; %)	Outcome Achieved (partially; %)	Outcome Achieved (fully + partially; %)
Project-level outcomes				
1.1 - Access to goods and services	19	63	11	74
1.1.1 - Access to goods and services (MSME)	1	0	0	0
1.1.2 - Access to goods and services (female)	0	n.a.	n.a.	n.a.
1.1.3 - Access to goods and services (customers)	4	100	0	100
1.1.4 - Access to goods and services (miscellaneous)	14	57	14	71
1.2 - Quality and affordability of goods and services	10	80	10	90
1.4 - Improved living standards (earnings) of individuals	2	0	50	50
1.6 - Economic return	1	0	0	0
3.1 - Increased employment	10	60	40	100
3.2 - Improved capacity and skills	1	100	0	100
4.1 - Increased transfers to the government	6	17	67	84
6.2 - Greenhouse gas reduction	6	33	0	33
6.3 - Efficient use of resources	3	67	33	100

(continued)

Outcome Type	Outcomes (no.)	Outcome Achieved (fully; %)	Outcome Achieved (partially; %)	Outcome Achieved (fully + partially; %)
7.2 - Induced or indirect employment	1	100	0	100
7.3 - Export sales	2	50	0	50
8.1 – Governance	1	0	0	0
Total project-level outcomes	62	55	21	76
Foreign investment–level outcomes				
9 - Business and sector practices	4	0	0	0
10 - Market development	9	33	44	77
11 - Development reach	1	0	0	0
12 - Sustainability	0	n.a.	n.a.	n.a.
13 - Signaling effects	2	100	0	100
Total foreign investment–level outcomes	16	31	25	56
Total outcomes	78	50	22	72

Source: Independent Evaluation Group.

Note: Of the total project-level outcomes, 10 percent were considered not achieved because the results could not be verified. Of the total foreign investment–level outcomes, 13 percent were considered not achieved because their results could not be verified. MIGA = Multilateral Investment Guarantee Agency; MSME = micro, small, and medium enterprise; n.a. = not applicable.

Monitoring and evaluation shortcomings prevented the validation of some development outcomes and may have depressed development outcome ratings. Of the 78 outcomes examined in the deep dive, 28 percent were not achieved either fully or partially, including 24 percent of project-level outcomes and 44 percent of foreign investment–level outcomes. Moreover, the Development Effectiveness Indicator System did not track 69 percent of outcomes. In many cases, MIGA or IEG used other information sources to validate the outcome claim. However, 10 percent of project-level outcomes

and 13 percent of foreign investment–level outcomes could not be verified by MIGA or IEG at all because of a lack of evidence, confirming the *RAP 2023* Concept Note's hypothesis that foreign investment–level outcomes are more challenging to measure than project-level outcomes.

Relationship between Outcomes and Project Performance Ratings

The *RAP* deep-dive analysis showed that projects that achieved more of their expected outcomes achieved higher development outcomes ratings. According to Project Evaluation Report guidelines, the development outcome ratings of MIGA guarantee projects are assigned at the project level and subdimension level and not at the outcome level. Therefore, this *RAP* expanded the outcome type analysis conducted by *RAP 2021* by examining the relationship between project outcomes and development outcome ratings. It generally found that projects with higher levels of outcome achievement had higher development outcome ratings. Projects with a successful development outcome rating achieved 80 percent of their outcomes, whereas projects with an unsuccessful rating achieved only 22 percent of their outcomes (table 4.3). A project's achievement of foreign investment–level outcomes was not as much of a determining factor for development outcome ratings' success, but this could be because of the very small sample size. Nevertheless, successful projects achieved 50 percent of their foreign investment–level outcomes, whereas unsuccessful projects achieved none of their foreign investment–level outcomes. Outcome achievement was also the main difference between borderline projects, with mostly successful projects achieving 69 percent of their outcomes but mostly unsuccessful projects only achieving 45 percent.

Table 4.3. MIGA Guarantee Project Development Outcome Ratings and Underlying Outcome Achievement Rates

Development Outcome Rating	Projects (no.)	Outcomes (no.)	Overall Outcome Weighted Achievement Rate (%)	Project-Level Outcome Weighted Achievement Rate (%)	Foreign Investment–Level Outcome Weighted Achievement Rate (%)	Project Business Performance Average Rating	Environmental and Social Effects Average Rating
Successful	5	25	80	89	50	3.0	3.0
Mostly successful	5	24	69	79	43	2.4	3.2
Mostly unsuccessful	4	20	45	44	50	2.0	2.8
Unsuccessful	2	9	22	25	0	1.0	3.0

Source: Independent Evaluation Group.

Note: Achievement rates at the project level are measured by efficacy of outcome achievements in projects, with the following weights: outcome achieved = 1, partly achieved = 0.5, not achieved = 0, and cannot verify = 0. Project business performance and environmental and social effects ratings' numerical values are as follows: excellent = 4, satisfactory = 3, partly unsatisfactory = 2, and unsatisfactory = 1. MIGA = Multilateral Investment Guarantee Agency.

[1] The ratings of Multilateral Investment Guarantee Agency (MIGA) guarantee projects are reported on a six-year rolling basis, given the small sample of the evaluated projects. The guarantee projects evaluated in fiscal years (FY)17–22 included 100 percent of the mature projects, with guarantee contracts issued during FY12–18.

[2] MIGA self-evaluations for one guarantee project in the FY20 program and seven guarantee projects in the FY22 program were received by the Independent Evaluation Group in the third quarter of FY23, whereas one was pending for one project in FY22. Therefore, Independent Evaluation Group validations of these projects have not been completed yet. In addition, MIGA self-evaluations for one project in FY21 and nine projects in FY22 have been deferred to the FY23 evaluation program due to sensitive circumstances.

[3] All 16 MIGA guarantee projects in the FY20–22 cohort covered in this *Results and Performance of the World Bank Group* deep-dive analysis of performance factors were evaluated and validated after March 2020.

5 | Conclusions and Future Directions

The purpose of this chapter is to provide concluding remarks and propose ways forward. This *RAP*'s analysis for each World Bank Group institution—the World Bank, IFC, and MIGA—consisted of different data, cohort sampling, terminologies, and methods of analysis. As such, the conclusions and future directions are presented separately for each institution.

World Bank

Conclusions

World Bank projects maintained or improved their performance ratings between FY21 and FY22 despite the challenges posed by COVID-19. This is the first *RAP* to include a substantial number of projects affected by the pandemic in its cohort; however, as most of these projects were already at an advanced stage of implementation when the pandemic began, the amount of their project life that was exposed to the pandemic was limited. Moreover, the presence of a sample selection bias, with more successful projects being overrepresented in this cohort, is a limitation that calls for a cautious interpretation of findings on the pandemic's impacts on projects. Because project cohorts in the next few years may exhibit a more accurate reflection of the extensive repercussions of the COVID-19 pandemic, it is likely that project performance rating trends will change downward in the future when more projects with prolonged exposure to COVID-19 are integrated into the rating trends.

World Bank projects were able to adapt to various implementation challenges, including COVID-19–related disruptions and other obstacles. The fact that overall project performance did not suffer is a testament to the resilience and adaptability of project teams. The low technical and organizational capacity of implementing agencies emerged as a key implementation

challenge, especially in projects that failed to adequately identify and mitigate institutional capacity risks. In addition, COVID-19 disrupted project implementation in various ways and led to increased project restructuring and adaptive management, which helped World Bank projects achieve desired results despite the many challenges. However, this *RAP* did not analyze in-depth the types of changes in results framework during restructurings—a potential area of analysis for future *RAP*s.

The World Bank's improved M&E quality documented project achievements and challenges, enabling adaptive management. Increased M&E quality, as reflected in M&E quality ratings, served as an effective early warning tool that enabled project adaptation and provided sufficient evidence of project achievements. Project indicators were well aligned with project objectives, and these adequate indicators contributed to improved efficacy ratings. Nevertheless, M&E challenges persist in measuring the achievement of enhanced capacity of institutions' objectives, which relied on intermediate outcome indicators to demonstrate the extent to which the project enhanced the capacity of institutions to perform their functions.

Future Directions

Strengthen project capacity to identify and mitigate risks during project preparation, especially the risk of low implementing agency capacity. Risk management by World Bank project teams and the technical capacity of implementing agencies were key factors in successful project implementation. Indeed, the weak capacity of implementing agencies emerged as the predominant underlying risk in projects that failed to adequately identify and mitigate risks. This underscores the need for World Bank project teams to conduct comprehensive risk assessments and develop robust mitigation strategies that prioritize capacity risks, especially in countries where local capacity limitations are common. This future direction aligns with the *RAP 2022* proposal to strengthen country programs' ability to assess implementation capacity risks.

Continue improving M&E as both an adaptation and accountability tool. The World Bank took a proactive approach to adapt and restructure projects as needed during the COVID-19 crisis by closely monitoring projects'

progress and identifying emerging challenges. M&E frameworks also provided sufficient evidence on project achievements. Thus, there is a valuable opportunity to scale up project monitoring, adaptation, and restructuring into postpandemic contexts and, more generally, beyond crisis scenarios. This will help maximize the resilience and performance of World Bank projects. Furthermore, there are still areas in which the World Bank can continue to improve the M&E frameworks for greater accountability. In particular, the World Bank could enhance its ability to measure institutional capacity outcomes in line with the World Bank's outcome orientation agenda. This future direction is consistent with *RAP 2021*, which shows that not all projects with institutional strengthening objectives have adequate indicators to measure them.

International Finance Corporation

Conclusions

IFC investment and advisory project performance ratings declined only slightly despite their exposure to COVID-19 and the more challenging operating environment. IFC's investment project development outcome success ratings decreased from 53 percent in CY19–21 to 50 percent in CY20–22. The development effectiveness success ratings of IFC advisory projects have been improving since FY15–17 but declined from 60 percent in FY19–21 to 54 percent in FY20–22. IFC's XPSR and Project Completion Report self-ratings also declined. Most IFC investment projects took place during some part of the COVID-19 pandemic, with, on average, 24 percent of their project lives occurring during COVID-19. The pandemic-related disruptions created a challenging operating environment for IFC investment projects. Although IFC restructures loan agreements and reschedules loan repayments, it has no formal procedures to modify investment projects' development objectives in response to a crisis.

Several factors besides COVID-19 negatively affected IFC's investment project performance, including economic issues, business risks, and market competition. Economic factors reduced demand for IFC client products and services and lowered the project companies' operational and financial performance compared with the projections at the Board approval stage.

Financial sector projects dealing with high business risks moved away from lending to riskier segments to preserve capital. In the real sector, adverse business factors related to cyclicality, a downturn in the markets, or untested and flawed business models affected investment project performance. Higher-than-expected competition led to investment projects missing operational targets and contributed to reduced operating margins and profitability. The private sector reacted quickly to the changing economic landscape and showed remarkable resilience and adaptability. Similarly, strong sponsor ability and technical expertise emerged as two of the most important factors to positively affect investment project performance.

IFC work quality and additionality ratings, which are associated with project development outcome ratings, declined in CY20–22. Overall, IFC work quality success ratings for investment projects declined from 60 percent in CY19–21 to 55 percent in CY20–22, while IFC additionality success ratings declined from 59 percent to 54 percent over the same period. IFC additionality success ratings in challenging markets such as Africa, IDA and blend, and FCS countries were lower than the IFC average. IFC particularly had difficulties in delivering nonfinancial additionality in these challenging markets.

IFC investment project objectives were highly outcome oriented, although outcome achievement rates were low. Overall, 100 percent of the 170 IFC investment projects reviewed in the *RAP*'s deep-dive analysis pursued project-level outcomes, whereas 74 percent pursued market-level outcomes. IFC investment projects fully achieved 45 percent of their total intended outcomes and partially achieved 22 percent. IFC investment projects achieved market-level outcomes at about the same rate as they achieved project-level outcomes. This contradicts the hypothesis in the *RAP 2023* Concept Note that market-level outcomes are much harder to achieve than project-level outcomes. This *RAP* shows that IFC investment projects that achieved more of their intended outcomes achieved higher development outcome ratings.

Some outcomes could not be verified because of a lack of appropriate result measurement indicators and evidence, depressing outcome achievement rates. Most of these investment projects in the *RAP 2023* cohort were not subject to an AIMM assessment at their approval and continued to be

monitored in the Development Outcome Tracking System. In many cases, IFC or IEG used other available information sources to validate project outcome claims. However, 8 percent of intended outcomes could not be verified because of a lack of evidence, potentially depressing IFC's investment project outcome achievement rates.

Future Directions

Improve the delivery of IFC additionality in difficult markets to enhance investment project outcomes. Difficult markets include those in FCS, Africa, and IDA and blend countries, in particular. We found that IFC additionality success ratings were particularly low in a large share of investment projects in these markets. The IFC 3.0 strategy aims to ramp up its investment program in these challenging markets. Higher realized IFC additionality in these challenging markets will make it more likely for IFC investment projects to achieve their objectives. IFC can add value to projects in these markets in several ways. For example, IFC delivers tailored financing but can also increase its provision of industry expertise and capacity-building advisory services, improve corporate governance, and enhance the environmental and social standards and practices of clients. Improving the delivery of IFC additionality would require IFC to adopt a proactive approach to ensure that additionality promises made at approval, particularly nonfinancial additionalities, are fulfilled and properly monitored during the investment project's life.

Further strengthen the selection of indicators and the measurement and tracking of intended development outcomes of investment projects. These measures would facilitate the monitoring of project development outcome progress and better reflect actual achievement. *RAP 2021* highlighted the challenges in measuring development outcomes, particularly at the market level, and this *RAP* showed that these challenges continue to be an issue. We found that monitoring data were not available for many intended development outcomes of IFC investment projects in the *RAP 2023* cohort. As such, IFC has an opportunity to improve its design and implementation of monitoring indicators to ensure that they can measure and track the achievement of intended project outcomes of investment

projects. This would require IFC to provide clear definitions and sources for chosen indicators and ensure that clients have the capacity to measure them. That said, the investment projects in the *RAP 2023* cohort predate the rollout of IFC's AIMM framework, which requires IFC to track all project claims until the AIMM target year, which could improve some of these monitoring issues. IFC confirmed that it has increased the use of standardized indicators, improved regular monitoring, and engaged in an ongoing effort to establish a new data platform for data tracking and reporting for investment projects approved under the AIMM system. Appropriate implementation of these measures could result in improvements in the measurement and tracking of intended outcomes, although IEG has not yet been able to systematically validate these claims as very few IFC investment projects approved under the AIMM framework have been evaluated so far.

Multilateral Investment Guarantee Agency

Conclusions

MIGA guarantee project development outcome ratings remained stable over the last six years but were slightly lower over the last three years, partially due to pandemic-related market challenges. The implementation of half of MIGA guarantee projects in the last three years was affected by COVID-19 to at least some extent. On average, these MIGA guarantee projects were exposed to the pandemic for 27 percent of their project lives. This suggests, and project evaluations confirmed, that FY20–22 projects operated under more challenging conditions during the pandemic. As such, MIGA's overall development outcome success rates on a six-year rolling basis remained stable in FY17–22 but were lower on a three-year rolling basis in FY20–22, which coincided with the onset of the pandemic and the more challenging operating environment. Like IFC, MIGA has no formal procedures for re-structuring development-related objectives or outcome targets during project implementation or crises.

Several factors besides COVID-19 negatively affected MIGA's project performance. Cost overruns and construction delays along with foreign exchange issues were the most common factors undermining project implementation.

Capable sponsors showed resilience and flexibility in helping MIGA guarantee projects adapt to the challenging economic landscape caused by COVID-19. For example, some MIGA hospital projects adjusted their services to assist the government in addressing new medical demands during the pandemic. Favorable legal and regulatory environments also helped projects effectively counter adverse factors.

MIGA's project objectives were highly outcome oriented, but their achievement was relatively low. As in IFC, approximately one out of five intended outcomes from MIGA projects were foreign investment–level outcomes, with the rest being project-level outcomes. However, MIGA was even more outcome oriented, with 86 percent of projects pursuing foreign investment–level outcomes. That said, foreign investment–level outcomes were more difficult for MIGA to achieve than project-level outcomes, as hypothesized in the *RAP 2023* Concept Note and demonstrated in *RAP 2021*. MIGA guarantee projects fully achieved 31 percent and partially achieved 25 percent of their foreign investment–level outcomes, compared with 55 percent full achievement and 21 percent partial achievement of their project-level outcomes. This had a direct effect on ratings; this *RAP* shows that MIGA projects that achieved more outcomes had higher development outcome ratings. Further undermining MIGA's outcome achievement were shortcomings in project M&E, particularly a lack of appropriate indicators to measure intended project outcomes. Subsequently, 69 percent of project outcomes were not tracked by MIGA. MIGA and IEG used supplementary data to separately verify many of these outcomes, but 10 percent of project-level outcomes and 13 percent of foreign investment–level outcomes still could not be verified by MIGA or IEG because of a lack of evidence. This may have affected MIGA's development outcome ratings.

Future Directions

Enhance project preparation work quality to strengthen the performance of MIGA guarantee projects. We found that MIGA work quality was rated lower than satisfactory in half of guarantee projects. MIGA could undertake more comprehensive project risk assessments, estimate detailed operational and financial projections with clear targets, and account for

stricter downside scenarios. These up-front actions would help project teams enhance the awareness or understanding of potential project risks, consider mitigation mechanisms, and set clear project expectations. In PPP projects, MIGA could identify foreseeable macroeconomic developments, such as local currency depreciations that can increase the government's financial obligations, and assess the risks from these developments, for example, whether the government is willing or able to pay the increased obligations to reduce their sustainability risks. According to MIGA, project risk assessments have recently improved, and the current IMPACT framework incentivizes project teams to mitigate risks to the extent possible. However, IEG has not been able to validate these claims because MIGA's guarantee projects approved under the IMPACT framework have not yet been subject to evaluation.

Strengthen measurement and tracking of intended development outcomes, particularly at the foreign investment level. These measures would facilitate monitoring of project development outcome progress, would better reflect actual achievement, and would be especially helpful for tracking the achievement of intended foreign investment–level outcomes. MIGA could accomplish this by better defining its project development objectives, selecting relevant indicators to measure outcomes, and establishing appropriate mechanisms to gather results evidence and development impact data. This suggestion is in line with findings of *RAP 2021*, which noted that many MIGA guarantee projects lacked sufficient evidence to rate project outcomes; however, MIGA's evidence collection has improved in recent years.

Bibliography

Bravo, Leonardo, Ariya Hagh, Roshin Joseph, Hiroaki Kambe, Yuan Xiang, and Jos Vaessen. 2023. *Machine Learning in Evaluative Synthesis: Lessons from Private Sector Evaluation in the World Bank Group*. IEG Methods and Evaluation Capacity Development Working Paper Series. Independent Evaluation Group. Washington, DC: World Bank.

Denizer, Cevdet, Daniel Kaufmann, and Aart Kraay. 2013. "Good Countries or Good Projects? Macro and Micro Correlates of World Bank Project Performance." *Journal of Development Economics* 105 (November): 288–302.

ECG (Evaluation Cooperation Group). 2011. *Good Practice Standards for the Evaluation of Private Sector Investment Operations*. 4th ed. Washington, DC: ECG.

IFC (International Finance Corporation). 2016. *IFC Strategy 3.0: IFC Strategic Vision—Making 60 Years of Experience Count*. Washington, DC: IFC.

IFC (International Finance Corporation). 2019. *Strategy and Business Outlook Update FY20–FY22: Gearing Up to Deliver IFC 3.0 at Scale*. Washington, DC: IFC.

IFC (International Finance Corporation). 2023. *Guidance: Guidance on Board Paper Template for Investment Staff*. Washington, DC: IFC.

Ika, Lavagnon A., Amadou Diallo, and Denis Thuillier. 2012. "Critical Success Factors for World Bank Projects: An Empirical Investigation." *International Journal of Project Management* 30 (1): 105–16.

Ortega Nieto, Daniel, Ariya Hagh, and Vivek Agarwal. 2022. "Delivery Challenges and Development Effectiveness: Assessing the Determinants of World Bank Project Success." Policy Research Working Paper 10144, 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.

World Bank. 2016. *Results and Performance of the World Bank Group 2016*. Independent Evaluation Group. Washington, DC: World Bank.

World Bank. 2017. *Guidelines for Reviewing World Bank Implementation Completion and Results Reports: A Manual for IEG ICR Reviewers*. Independent Evaluation Group. Washington, DC: World Bank.

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

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

World Bank. 2022a. *Results and Performance of the World Bank Group 2022*. Independent Evaluation Group. Washington, DC: World Bank.

World Bank. 2022b. *The World Bank's Early Support to Addressing COVID-19: Health and Social Response—An Early-Stage Evaluation*. Independent Evaluation Group. Washington, DC: World Bank.

World Bank. 2023a. *Global Economic Prospects*, June 2023. Washington, DC: World Bank.

World Bank. 2023b. *International Finance Corporation Additionality in Middle-Income Countries*. Independent Evaluation Group. Washington, DC: World Bank.

APPENDIXES

Independent Evaluation Group

Results and Performance of the World Bank Group 2023

Appendix A. Methodological Approach

As in past years, the 2023 *Results and Performance of the World Bank Group* (*RAP*) synthesizes Independent Evaluation Group (IEG) ratings and other evidence from World Bank Group self-evaluations and IEG validations and evaluations to give an aggregated picture of the results and performance of the World Bank, International Finance Corporation (IFC), and Multilateral Investment Guarantee Agency (MIGA).

The report includes a traditional update of performance ratings and in-depth analyses to explain the variations in performance and respond to the following evaluative questions:

1. How did IEG ratings change over time at the project and country levels across the various Bank Group institutions?
2. What has been the evolution of development outcomes pursued, measured, and achieved at the project level, and what is the relationship between outcomes and project performance ratings?
3. What factors affected Bank Group projects' implementation and performance in the COVID-19 pandemic context?

Three principles underpin the methodological approach of this *RAP*: continuity, innovation, and symmetry. The report embraces continuity by building on previous work while incorporating innovative elements in a balanced and consistent manner across different institutions within the Bank Group, accounting for their differences in evaluation and rating in the symmetrical application of these approaches.

Methodology Approach for the World Bank Analyses

The Independent Evaluation Group's Evaluation Methodology for World Bank Lending Operations

The Implementation Completion and Results Reports (ICRs) prepared by World Bank staff are essential self-evaluation tools to assess projects and operations. As part of its validation work, IEG conducts independent reviews of the ICRs, known as Implementation Completion and Results Report Reviews (ICRRs). These reviews critically validate the evidence, results, and ratings of the ICRs, aligning them with the project's design documents as necessary. IEG also conducts fieldwork and collects additional data to assess 20–25 percent of the World Bank's lending operations annually through Project Performance Assessment Reports.

Project performance ratings for World Bank projects are derived from an objective-based methodology that, together with performance rating scales and criteria, was agreed on with Operations Policy and Country Services. Ratings are rubrics for assessing performance relative to a project or program's objectives. Ratings divide the World Bank's self-evaluation and IEG's validation narratives into categories or values that enable aggregation across operations.

Main Performance Ratings in World Bank Investment Operations

- Outcome refers to the extent to which a project efficiently achieved, or was expected to achieve, its relevant objectives. The outcome rating brings together three underlying dimensions: relevance, efficacy (objectives achievement), and efficiency. It is rated on a six-point scale: highly satisfactory, satisfactory, moderately satisfactory, moderately unsatisfactory, unsatisfactory, and highly unsatisfactory.
- Relevance is defined as the extent to which a project's objectives are consistent with current World Bank country strategies at the time of project closing. It is rated on a four-point scale: high, substantial, modest, and negligible.

- Efficacy is defined as the extent to which a project achieves, or was expected to achieve, its objectives, taking into account the objective's relative importance. The project's achievement of each individual objective is assessed based on the concept of "plausible causality." Efficacy ratings also reflect an assessment of the results framework's validity and use complementary data and evidence on the achievement of intended results. Both the efficacy of each individual objective and an overall efficacy in achieving the project development objective are rated on a four-point scale: high, substantial, modest, and negligible.
- Efficiency is a measure of how economic resources and inputs are converted to results. It indicates whether the costs involved in achieving project objectives were reasonable in comparison with the benefits and recognized norms (value for money). It is rated on a four-point scale: high, substantial, modest, and negligible.
- Bank performance refers to the extent to which World Bank services ensured quality project design and supported effective implementation through appropriate supervision in the achievement of development outcomes. Bank performance and its two constituent elements—quality at entry and quality of supervision—are rated on a six-point scale: highly satisfactory, satisfactory, moderately satisfactory, moderately unsatisfactory, unsatisfactory, and highly unsatisfactory.
- Monitoring and evaluation quality refers to the quality of the design and implementation of the monitoring and evaluation arrangements of the project and the extent to which the results are used to improve performance. Monitoring and evaluation quality is assessed at the project level and comprises monitoring and evaluation design, implementation, and use. It is rated on a four-point scale: high, substantial, modest, and negligible.

The evaluation methodology for development policy financing projects changed in mid-2020. Table A.1 compares the old and new methodologies of the ICRR.

Table A.1. Old and New ICRR Methodology for Development Policy Financing

Area	Old Methodology	New Methodology
Relevance of objective and design	» Relevance of objectives (rated) » Relevance of design (rated)	» Relevance of objectives (not rated) » Relevance of prior actions (rated)
Relevance of results indicators	» Not routinely discussed, not rated	» Discussed and rated
Achievement of objectives (efficacy)	» Discussed and rated based on achievement of targets	» Discussed and rated based on achievement of relevant results indicator targets (or other evidence)
Outcome	Rated based on the following: » Relevance of objectives and design » Achievement of objectives (efficacy)	Rated based on the following: » Relevance of prior actions » Achievement of objectives (efficacy)
Risk to development outcome	» Discussed and rated	» Discussed but not rated
Bank performance	Overall rating based on the following: » Quality at entry (rated) » Quality of supervision (rated)	Overall rating based on the following: » Design (rated) » Implementation (rated)
Borrower performance	» Rated	» Dropped
Monitoring and evaluation design, implementation, and use	» Rated	» Dropped

Source: Independent Evaluation Group.

Note: ICRR = Implementation Completion and Results Report Review.

Analysis, Data Sources, and Sample Coverage

Table A.2 indicates the data sources and sample coverage used in this *RAP*.

Table A.2. *RAP 2023* Data Sources and Sample Coverage

Analysis	Data Sources	Sample Coverage
World Bank project performance ratings	IEG data on ICRR or Project Performance Assessment Report ratings	World Bank lending projects closed during FY12–22 and evaluated by the IEG as of June 30, 2023
Analysis of World Bank project outcome types	IEG ICRR rating data and documents, World Bank Data Explorer data	*RAP 2023* portfolio of 273 World Bank investment project financing projects closed since March 2020 and evaluated by IEG as of December 2022
Analysis of World Bank project results frameworks	IEG ICRR rating data and documents, World Bank Data Explorer data, World Bank SAP data	*RAP 2023* portfolio (as above)
Factors affecting World Bank project implementation	World Bank project Implementation Completion and Results Reports	*RAP 2023* portfolio (as above); and prepandemic portfolio of 398 investment project financing projects (i) closed between July 2017 and February 2020 and (ii) with ICR completed before March 2020 and (iii) and evaluated by IEG as of December 2022
World Bank project restructuring patterns (including split)	IEG ICRR data, World Bank Data Explorer data	*RAP 2023* portfolio (as above). Prepandemic portfolio (as above)

Source: Independent Evaluation Group.

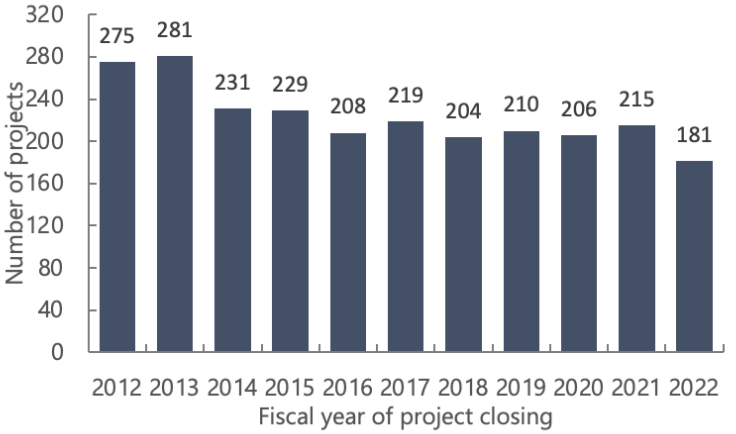
Note: FY = fiscal year; IEG = Independent Evaluation Group; ICRR = Implementation Completion and Results Report Review; *RAP* = *Results and Performance of the World Bank Group*.

This *RAP*'s analysis of performance rating trends covers 2,973 World Bank lending projects that closed between fiscal year (FY)12 and FY22 and were evaluated by IEG through either an ICRR or a Project Performance Assessment Report as of June 30, 2023. IEG evaluated 96 percent (2,459 of

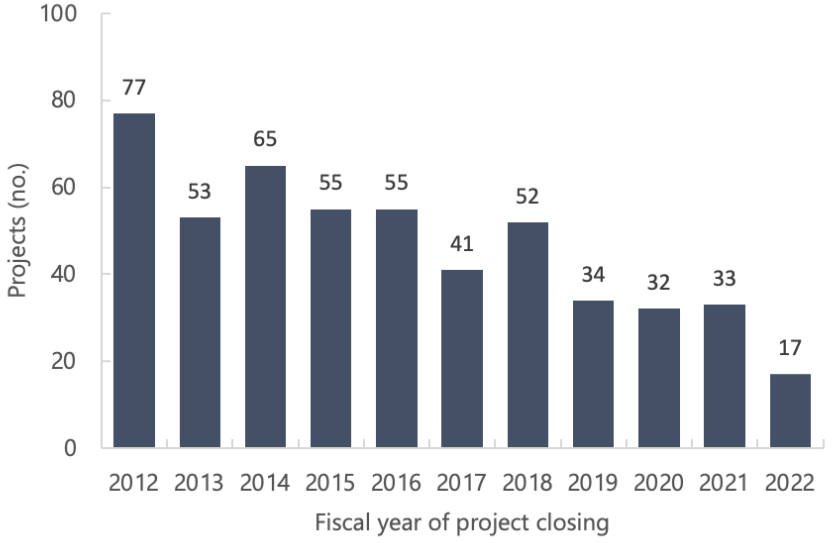
2,552) of the investment project financing and Program-for-Results projects closed during FY18–22. As of June 30, 2023, IEG's pipeline had 113 ICRRs for all lending types, approximately 19 percent of which had been in the pipeline for 180 days or longer (figure A.1).

Figure A.1. Coverage of World Bank Project Ratings Data

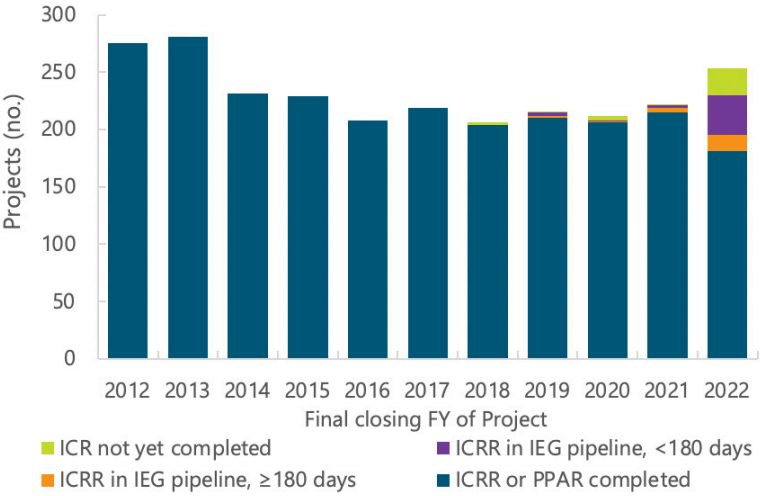
a. Investment project financing and Program-for-Results Projects



b. Development policy financing Projects



c. Coverage of World Bank investment project financing and Program-for-Results projects ratings data



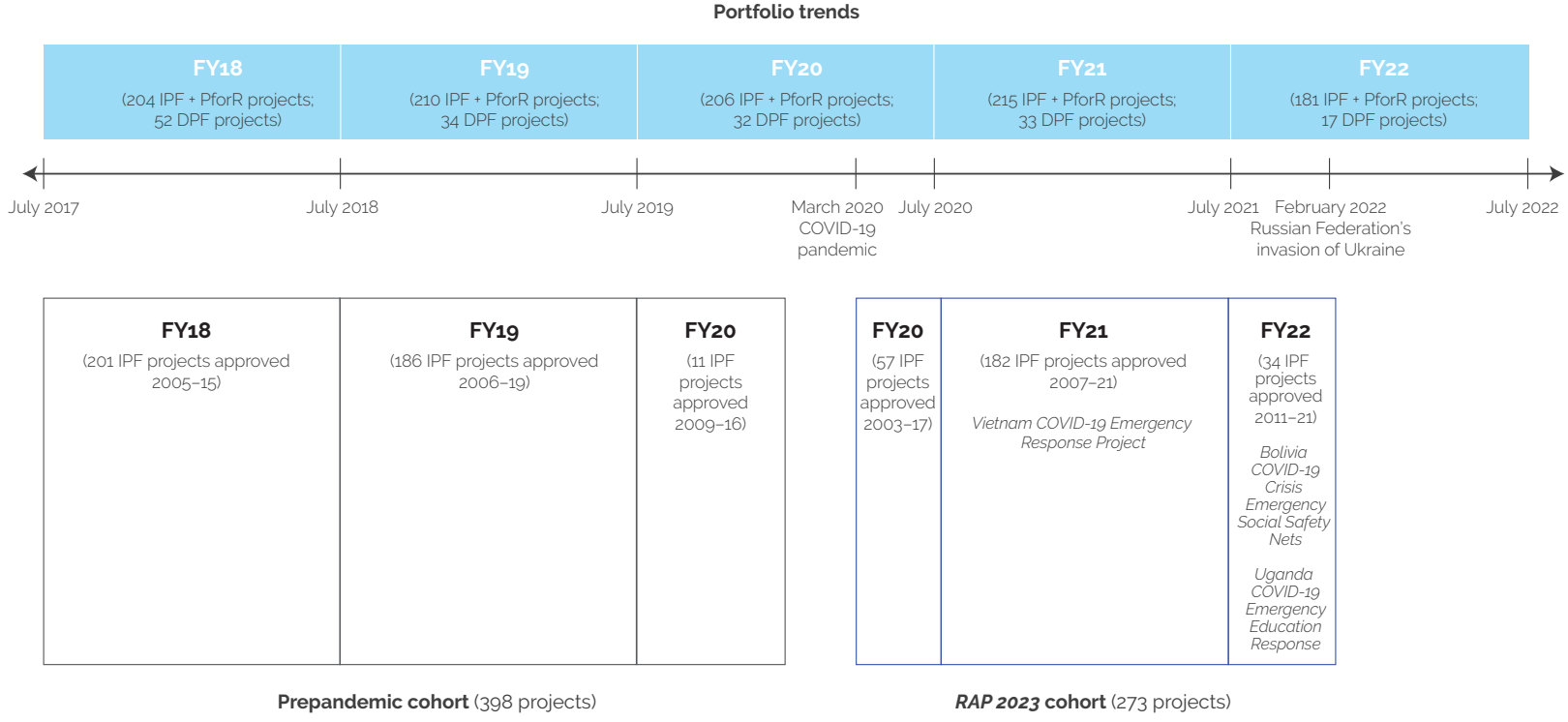
Source: Independent Evaluation Group..

Note: Data as of June 30, 2023. For fiscal years 2012–17, panel c shows only completed ICRRs and PPARs. For each project, the graph includes the latest IEG project evaluation (ICRR or PPAR). ICRR = Implementation Completion and Results Report Review; IEG = Independent Evaluation Group; PPAR = Project Performance Assessment Report.

Project Exposure to the COVID-19 Pandemic and Sample Selection Bias

This is the first *RAP* with a substantial number of projects affected by the COVID-19 pandemic. The update of the rating trends for FY20–22 includes 684 lending operations that were closed and evaluated by IEG as of June 30, 2023. The in-depth analyses presented in this *RAP* focused on 273 investment project financing projects that were closed between March 2020 and April 2022 and evaluated by IEG as of December 2022. Figure A.2 illustrates the difference in scope between the rating trends, the *RAP 2023* cohort, and the prepandemic cohort. In addition, projects in the *RAP 2023* cohort had limited exposure to the COVID-19 pandemic (box A.1).

Figure A.2. Composition of the Overall Portfolio in Rating Trends, the Prepandemic Cohort, and the *RAP 2023* Cohort



Source: Independent Evaluation Group.

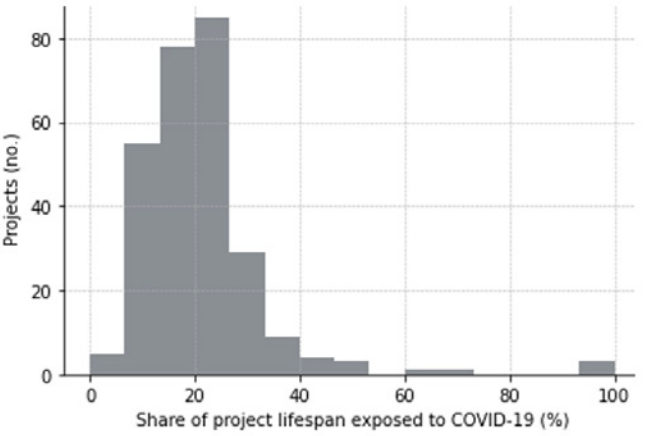
Note: FY = fiscal year; IPF = investment project financing; DPF = development policy financing; PforR = Program-for-Results; *RAP* = *Results and Performance of the World Bank Group*.

Box A.1. Project Exposure Time to COVID-19

Projects in the *Results and Performance of the World Bank Group* 2023 cohort were at an advanced stage of implementation when the pandemic struck. On average, they were exposed to COVID-19 for 14 percent of their total lifetime. This measurement is calculated by dividing the project life span after February 2020 by the project's overall duration. For a small number of still-active projects with completed Implementation Completion and Results Report Reviews, the exposure measure is calculated by dividing the period after February 2020 until the Implementation Completion and Results Report completion date by the project's overall duration. As shown in the histogram in figure BA.1.1, the project exposure to COVID-19 exhibits a right skew. Approximately half of the projects experienced an exposure level below 12 percent of the project life, and 90 percent had an exposure level below 23 percent.

In fact, some projects reported that the COVID-19 pandemic had limited impact on the quality, nature, or extent of implementation because they were already nearing completion. In some cases, the pandemic even allowed some tasks to become more efficient, emphasizing the importance of connectivity and digitalization in navigating, responding to, and recovering from such crises. Some projects adapted to the pandemic's constraints using online platforms and remote training services. The pandemic also led to more efficient processes and accelerated uptake of e-services. Implementation was not significantly affected in some cases, thanks to swift reactions and support of remote work.

Figure BA.1.1. Project Exposure Time to COVID-19



Source: Independent Evaluation Group.

Limitations of data coverage. Performance ratings in FY22 are likely to be overrepresented by more successful projects, thus suffering a sample selection bias. The term *selection bias* refers to a distortion in the rating data of the included projects, stemming from the limited representativeness of the entire population of projects that closed within a specific fiscal year simply because the rest of the projects have not been evaluated yet. Each year, as more ICRRs are completed, they are accounted for in the performance ratings of the fiscal year in which those projects were closed rather than the fiscal year in which the evaluation took place. Projects with ICRs and ICRRs completed relatively quickly after project closure tend to have higher ratings than projects with delayed ICRs and ICRRs, as shown by the negative association between project performance ratings and time elapsed between project closure and completion of the ICRR (figure A.3). This pattern has also been observed in previous years. Every year, the update of rating trends results in a slight decline in performance ratings from previous *RAP* reports. Implementation Status and Results Report ratings confirm this pattern. As of June 29, 2023, ICRs were still pending for 20 projects that closed in FY22; ICRRs were in process for 49 projects. Inspection of the latest Implementation Status and Results Report ratings of progress on achievement of project development objectives (PDOs) in FY22 shows that projects with completed ICRRs—that were therefore included in the rating trends—have a higher average Implementation Status and Results Report rating than those with in-process ICRRs and higher still than projects with uncompleted ICRs (table FA.3.1). Rating trends should be interpreted carefully because they are likely to decrease. This is especially true as more projects with extended exposure time to COVID-19 are incorporated into project rating trends. The *RAP 2023* cohort is even more likely to be overrepresented by projects exhibiting higher ratings than the portfolio included in the rating trend because of the December 2022 cutoff point set to accommodate the time required for data collection and in-depth analysis. As such, the *RAP 2023* cohort may not accurately reflect the broader population of projects affected by the COVID-19 pandemic, potentially skewing the findings toward better ratings.

Figure A.3. Sample Selection Bias: Overrepresentation of More Successful Projects

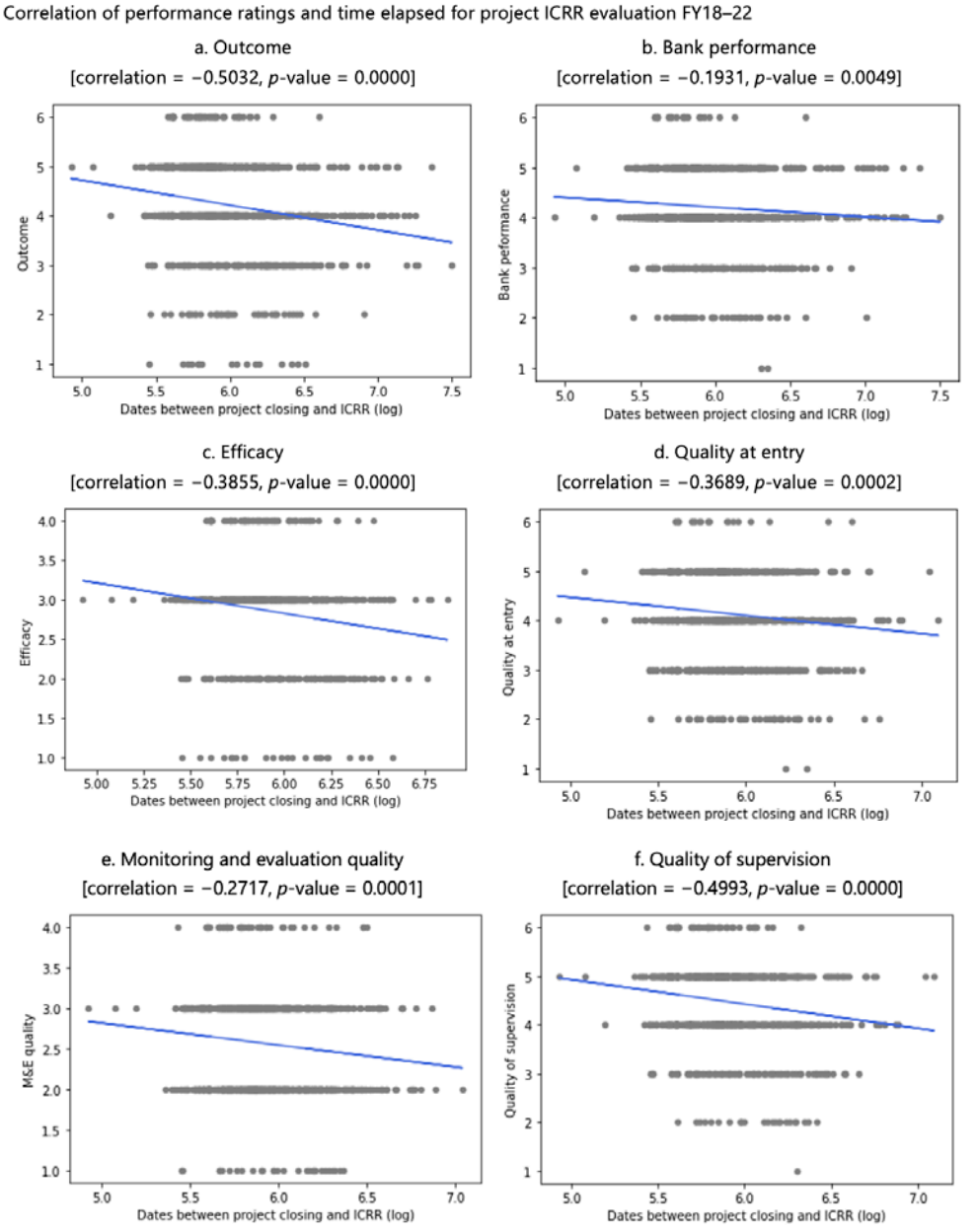


Table FA.3.1. Average Latest ISR Rating on Progress toward Achievement of Project Development Objective According to Evaluation Status in FY22

Evaluation status	Projects (no.)	Average ISR rating
ICR not submitted	20	4.37
ICRR IEG pending (180 days or more)	16	4.44
ICRR IEG pending (less than 180 days)	33	4.52
ICRR completed	170	4.56

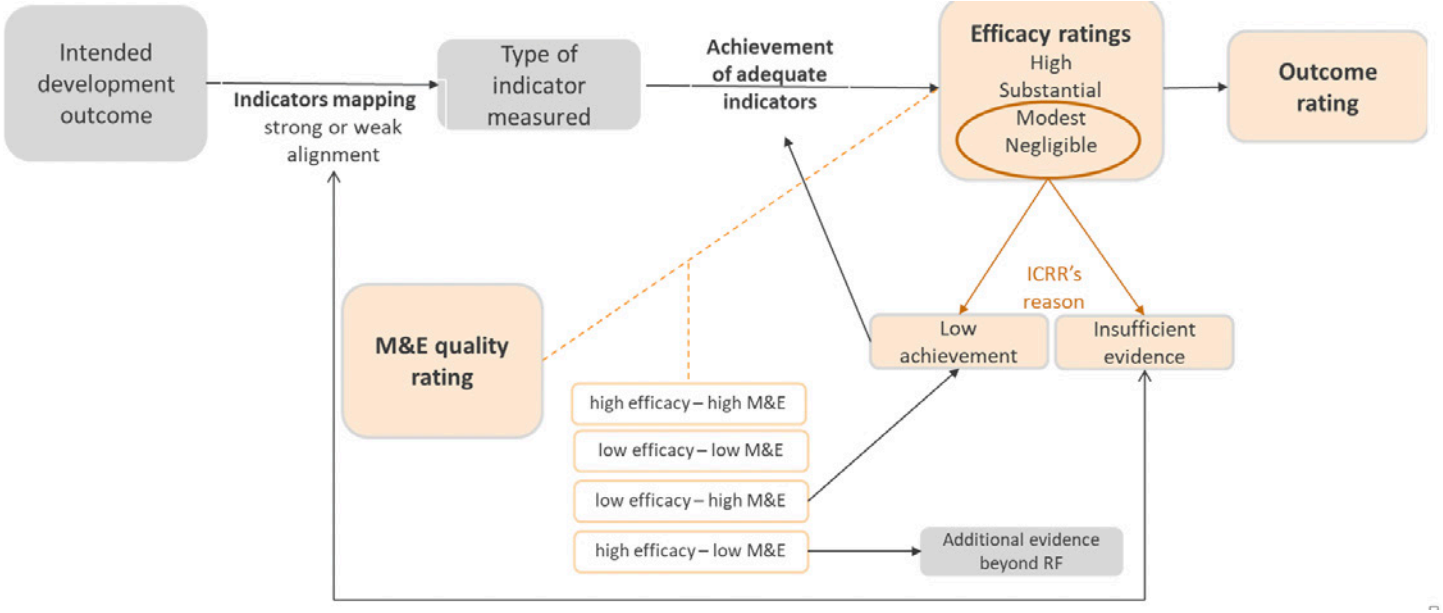
Source: Independent Evaluation Group.

Note: M&E = monitoring and evaluation; ICR = Implementation Completion and Results Report; ICRR = Implementation Completion and Results Report Review; IEG = Independent Evaluation Group; ISR = Implementation Status and Results Reports. ISR rates the progress towards achievement of PDO on a six-point scale.

Analysis of Project Development Outcomes and the Results Framework

This *RAP* examined the relationship between intended outcomes, measured outcomes, and key performance ratings for World Bank investment operations (figure A.4).

Figure A.4. Links among Types of Intended, Measured, and Achieved Outcomes; Monitoring and Evaluation Quality; and Efficacy Ratings



Source: Independent Evaluation Group.

Note: ICRR = Implementation Completion and Results Report Review; M&E = monitoring and evaluation; RF = results framework.

Outcome typology. To identify the development outcomes underlying project performance ratings for World Bank investment operations, we used the outcome typology developed by *RAP 2021* (World Bank 2021). This outcome typology was designed to capture the type of change envisioned by project objectives and consisted of 16 outcome types derived from typical project theories of change and select corporate objectives (figure A.5).

Figure A.5. *RAP 2021* Outcome Typology

Source: Independent Evaluation Group.

Note: RAP = Results and Performance of the World Bank Group.

Data sources. For the analysis of outcome types, we used an ICRR data set of 695 individual objectives from the 273 investment project financing projects included in the *RAP 2023* cohort. The data set includes both original and revised objectives and a description of the theory of change underlying each objective. At the ICRR stage, an experienced IEG evaluator parses the PDO statement into its separate individual objectives and assigns corresponding efficacy ratings. The evaluators' skills and expertise ensured that the individual project objectives were logical and well defined.

For the analysis of results frameworks indicators, we used a data set of 4,808 indicators of the 273 investment project financing projects included in the *RAP 2023* cohort, of which 1,458 were PDO indicators (30 percent) and 3,350 were intermediate results indicators (70 percent). PDO indicators are designed to measure the intended impact or outcome of the project and are usually limited in number. Intermediate results indicators, instead, are designed to measure the progress of specific components or activities within the project that contribute to the achievement of the project's overall goal, and there tend to be more of them than PDO indicators.

Coding process. The coding team comprised nine expert evaluators who were assigned 20 to 40 projects each based on distribution by Global Practice and their area of evaluation expertise. The coding process of classifying the development objectives and results frameworks indicators according to the outcome typology entailed different stages:

1. Training sessions and guiding material. The team held three training sessions for the coders to familiarize themselves with the typology, taking advantage of supplementary materials from *RAP 2021*, including handouts and definitions, and previously coded data.

2. Coding software and template. The team used an MS Excel template for manual coding purposes. Each individual objective was assigned to a maximum of three outcome types depending on the statement of the individual objective and its corresponding theory of change text. For example, an objective to "restore and expand access to reliable electricity services" was classified as quality of services improved and access to services expanded. Most individual objectives (57 percent) were linked to a single development outcome type, while 32 percent were assigned to two outcome types, and only 11 percent of individual objectives were associated with three outcome types. Similarly, another MS Excel template was developed to manually code all projects' indicators, that is, PDO indicators and intermediate results indicators. The team assessed indicators according to three dimensions:

   - Indicator outcome type. Similar to the individual objectives coding, each indicator was assigned to a maximum of three outcome types de-

pending on the nature and language of the indicator. This allowed the team to compare the outcome types of indicators with outcome types of the intended development outcomes in the objectives. In addition, each indicator was mapped to one or more objectives according to the ICRR Efficacy section.

» Indicator level. Each indicator was classified according to four levels: output, intermediate outcome, outcome, and high outcome. For instance, indicators that demonstrated the completion of project activities were classified as outputs. Examples of high outcome indicators include student learning gains, annual avoided carbon dioxide emissions, cholera case fatality rate, and reduction in fatalities and serious injuries per 100 million vehicle-km. The achievement of individual objectives, on average, was measured by PDO indicators, which predominantly focused on intermediate outcome indicators (46 percent), and outcome indicators (40 percent), while a few PDO indicators measured outputs (12 percent), and a mere 2 percent measured high outcomes. Instead, project intermediate results indicators, which aim at assessing the progress of specific project components or activities and offer operational insights into the project's progress, were mostly focused on lower-level indicators. They were mostly measuring outputs (53.6 percent), followed by intermediate outcomes (38.2 percent), outcomes (8.1 percent), and high outcomes (0.1 percent).

» Indicator adequacy. Each indicator was classified according to three levels: fully adequate, partially adequate, and not adequate. The assessment was based on the extent to which the indicator contributed to providing evidence on the achievement of the individual objective. Given the distinction between PDO and intermediate results indicators, not all results framework indicators are sufficient to demonstrate the achievement of the development objectives, yet they are necessary to provide evidence of the completion of project activities in line with the project's theory of change. For example, the individual objective of a transport project aimed "to reestablish lasting road access between provincial capital, district, and territories in the project impact area." The intermediate results indicator of the number of "condoms distributed" was classified

as not adequate because it did not provide evidence of the increased access to or quality of road conditions. The intermediate results indicator related to the "action plan to develop the road construction industry implemented" was classified as partially adequate because it contributed to providing evidence on the achievement of the individual objective to some extent. The intermediate results indicators of "reopened project roads in good to fair condition" and "roads in good and fair condition as a share of total classified roads" were classified as fully adequate because they measured the underlying intended individual objective. We found that 85 percent of development objectives had at least one fully adequate indicator to measure the intended development outcomes. On average, individual objectives had 65 percent fully adequate, 35 percent partially adequate, and zero inadequate PDO indicators.

3. Quality assurance. The coding process began with a pilot exercise involving three projects per coder, and pilot outputs were validated to ensure homogeneous understanding of the outcome typology and the assessment of indicators' adequacy across coders. To ensure intercoder reliability, the coding team conducted several rounds of reviews. At the midpoint of the coding process, coders with similar sectoral expertise were paired together to review a randomly selected batch of projects from each other's sector projects. They provided feedback to ensure the output's consistency, reliability, and validity, and identified and addressed areas for potential recoding or deletion. Progress was monitored using a tracking tool, and the evaluation team leader and other team members reviewed the coding for the entire portfolio after consolidating data from individual coders to produce the final coding output.

Limitations of the analysis of outcome types and indicators. The data set of indicators captures only the most recent version of results frameworks, which means it offers limited insight into changes made to frameworks throughout implementation. In addition, any indicators removed during the project cycle were excluded from the data set, further restricting the information available.

Analysis of Factors Affecting Projects' Implementation

We adopted a hybrid approach that combined inductive and deductive methods to identify the factors that affected the implementation of evaluated projects in FY20–22.

Taxonomy of factors affecting implementation. We used a slightly modified version of the World Bank's DeCODE (Delivery Challenges in Operations for Development Effectiveness) taxonomy,[1] relying on nearly all the preexisting category and subcategory definitions as originally developed (see table A.3 for a detailed description of the original taxonomy). The following adjustments were introduced:

1. Granularity of the taxonomy. Given the sparsity of observations for certain subcategories, as evidenced by the original DeCODE training set,[2] some factors were coded at the category level only: "coordination and engagement," "commitment and leadership," "legislation and regulation," "conflict and instability," "social and cultural," "environmental and geography," "business environment," "macroeconomic environment."

2. Adding sentiment. The term *sentiment* refers to the characterization of the language used in the ICR to describe a specific factor as either positively or negatively affecting the implementation of projects. Unlike the DeCODE taxonomy, which focuses on delivery challenges, we classified relevant extracts of text as factors that positively or negatively affected a project's implementation.

3. Renaming categories. The original category "overambitious objectives" was renamed "objectives (or design) scope" to avoid a negative connotation in cases in which project objectives or design were reported to be realistic and adequate.

4. Adding categories. Two categories not covered in the original taxonomy were added:

    - "Risk identification and mitigation" refers to appropriate (or lack of) up-front risk analysis or risk-mitigation actions. For instance, ICR text saying that "the risks were underestimated at the design stage, and the readiness to implement the project was overestimated" was classified

under this category as having a negative sentiment, whereas ICR text saying that "the project team appropriately identified key risks that could affect implementation and outcome and outlined mitigation measures to address them" was classified under this category as having a positive sentiment.

- "Adaptive management" refers to project course corrections during implementation to adapt and respond to shocks and unforeseen circumstances. Unlike risk identification and mitigation, which is an ex ante concept addressed during project preparation, adaptive management can be seen as an ex post concept. Examples include technical failures overlooked in the early stage that needed capacity building or flexibility built into World Bank operational policies and approaches that were realized later, such as recalibration of safeguards related to financing for vaccines.

5. Merging subcategories. After the coding process was finished, two subcategories, "skilled human resources" and "organizational capacity," were merged into a single subcategory named "skilled human resources and organizational capacity" because of their high similarity. (See the Supervised Machine Learning section.)

As a result of these changes, the revised taxonomy continued to be organized under three clusters (stakeholders, context, project) and comprised 17 categories, 7 of which also incorporated subcategories (table A.4).

Table A.3. Delivery Challenges in Operations for Development Effectiveness Taxonomy

Cluster	Category	Category Definition	Subcategory	Subcategory Definition
Stakeholders				
	Coordination and Engagement	Delivery challenges stemming from difficulty in coordination and engagement among stakeholders due to issues of administrative/ bureaucratic structure, unclear definition of roles, or inadequate engagement and communication strategies.	» Roles and Responsibilities	» Challenges that emerge when roles and responsibilities of different stakeholders are not clearly defined.
			» Stakeholder Engagement	» Challenges stemming from failure to adequately and actively engage beneficiaries or relevant stakeholders.
			» Awareness and Communication Strategy	» Challenges stemming from inability to raise awareness or unwillingness/inability to share relevant information with beneficiaries and/or the general public.
			» Bureaucratic Structure	» Administrative barriers or bureaucratic structures that impede and/or slow down coordination or engagements.
			» Inter- and Intragovernmental Relations	» Challenges caused by the difficulty of coordinating among different levels and structures of government with differing priorities and/or mismatches of resources, responsibilities, and/or expectations.

(continued)

Cluster	Category	Category Definition	Subcategory	Subcategory Definition
Stakeholders				
	Commitment and Leadership	Delivery challenges stemming from a change in leadership, shifts in priorities, or the absence of shared commitment and consensus among stakeholders.	» Change in Leadership and Administration » Opposition or Lack of Consensus » Change in Priorities or Lack of Commitment	» Challenges caused by leadership change in the government or relevant stakeholders. » Inability to find a solution that is acceptable to all major stakeholders, or opposition from stakeholder groups or individuals to a proposed intervention. » Issues caused by sudden changes in organizational priorities or the degree of commitment to a particular intervention.
	Human Resources and Organizational Capacity	Delivery challenges faced because of constraints caused by lack of skilled human resources, difficulties in acquiring necessary skills, or limited organizational capacity.	» Skilled Human Resources » Skill Transfer » Staff Turnover » Organizational Capacity	» Challenges caused by lack of appropriately skilled project staff. » Challenges caused by difficulty of imparting or acquiring new skills needed. » Challenges caused by short tenure of staff on projects. » Challenges caused by inability of an organization to execute interventions due to its overall institutional arrangements.

(continued)

Cluster	Category	Category Definition	Subcategory	Subcategory Definition
Context				
	Legislation and Regulations	Delivery challenges stemming from an unsupportive legal environment caused by lack of appropriate legal/regulatory framework, inordinate delays in promulgating laws, or complicated and time-consuming regulatory processes.	» Lack of Regulation and Legislation » Unsupportive Legal and Regulatory Process	» Challenges stemming from lack of or inadequate laws, regulations, or an appropriate legal framework. » Challenges that result from excessive and complicated legal or regulatory processes.
	Governance and Politics	Delivery challenges faced because of elections, opaque governance environment characterized by poor accountability, weak rule of law, political manipulation of projects, or corruption.	» Voice and Accountability » Corruption and Patronage » Political Interference » Electoral Cycles » Rule of Law	» Challenges caused by the inability of citizens to actively express their opinions and/or insufficient mechanisms to ensure transparency and hold service providers accountable. » Challenges stemming from the abuse of public power for private gain and/or favoritism toward patrons/clients/associates. » Challenges caused by steering decisions or projects for political purposes. » Challenges caused by elections and electoral processes. » Challenges caused by stakeholders not abiding by the rules and/or issues with contract or regulation enforcement, including judiciary problems.

(continued)

Cluster	Category	Category Definition	Subcategory	Subcategory Definition
Context				
	Conflict and Instability	Delivery challenges faced because of disruptions stemming from a conflict/postconflict situation, insecurity, or civil unrest.	» Crime and Violence » Civil Unrest and Armed Conflict » Postconflict Situation	» Challenges stemming from criminal violence and insecurity. » Challenges caused by protests, contentious mobilization, disputes, or active conflict within a country. » Challenges stemming from instability after armed conflict.
	Social and Cultural	Delivery challenges stemming from language barriers, social or cultural norms and practices, including gender and religion.	» Gender » Language » Culture, Religion and Ethnicity	» Challenges related to gender issues, discrimination, or disagreement over appropriate gender roles. » Difficulties caused by language barriers with partners or beneficiaries, or issues with linguistic discrimination. » Challenges caused by prevailing group practices or accepted social norms.
	Environment and Geography	Delivery challenges faced because of environmental characteristics, or difficulty accessing areas or populations.	» Geographic Access » Ecosystem	» Challenges stemming from problems accessing populations due to geographical barriers and remoteness. » Challenges specific to the ecological makeup of an area.

(continued)

Cluster	Category	Category Definition	Subcategory	Subcategory Definition
Context				
	Basic Infrastructure	Delivery challenges caused by constraints on power infrastructure, or insufficient communications or transportation systems.	» Information and Communication Technology » Energy and Electricity » Transportation	» Challenges stemming from deficiencies or mismatches in ICT. » Challenges caused by constraints on implementation because of lack of energy and electricity supply. » Challenges stemming from underdeveloped transportation systems and logistical networks.
	Disasters and Emergency Response	Delivery challenges caused by natural/man-made disasters or other unexpected emergency situations.	» Natural Disasters » Man-made Disasters » Epidemics	» Challenges stemming from natural disasters. » Challenges stemming from man-made disasters. » Challenges stemming from disruptions caused by epidemics.
	Business Environment	Delivery challenges caused by a weak private sector, or weak sector regulations.	» Private Sector Regulation » Weak Private Sector » Informal and Illegal Markets	» Challenges caused by the absence of regulations, or restrictive regulations, in the private sector. » Challenges stemming from the insufficient volume and/or lack of service delivery capacity of private sector entities, or the unestablished situation of the overall private sector. » Challenges caused by distortions of high informality and shadow/parallel markets.

(continued)

Cluster	Category	Category Definition	Subcategory	Subcategory Definition
Context				
	Macroeconomic Environment	Delivery challenges caused by instability, volatility, or interruptions in trade, market conditions, or financial systems.	» Trade Barriers	» Challenges caused by international or domestic restrictions on cross-border exchange of goods or services.
			» Financial Instability	» Challenges stemming from disruptions in the financial system.
			» Market Deterioration	» Challenges stemming from the shrinking of market size, or the price anomalies/distortions caused by systematic market failures.
			» Forex Volatility	» Challenges caused by sudden currency devaluation/depreciation or restrictions relating to transfer of forex.
Project				
	Project Design	Delivery challenges stemming from flaws in project design, including overly complicated design, overambitious objectives, inappropriate time allocation, or issues in identifying and selecting/targeting stakeholders and beneficiaries.	» Overambitious Objectives	» Challenges caused by setting targets that are unrealistically ambitious, or making the project design overly complex.
			» Time Allocation or Task Sequencing	» Challenges related to insufficient/excessive duration of a component, or inappropriate timing and sequence of task.
			» Stakeholder Selection	» Challenges caused by problems identifying/selecting appropriate stakeholders to engage.
			» Beneficiary Targeting	» Challenges with ensuring that the appropriate beneficiary group is targeted.

(continued)

Cluster	Category	Category Definition	Subcategory	Subcategory Definition
Project				
	Project Finance	Delivery challenges related to procurement, or fiduciary arrangements such as planning and budgeting, financing mechanisms, financial reporting, and auditing.	» Procurement	» Challenges caused by issues with procurement management systems, including poor contract management and delays.
			» Financing Mechanism	» Challenges related to the choice of financing mechanism or instrument.
			» Budgeting	» Challenges related to insufficient/inappropriate budget allocation, or caused by complex budget processes and management.
			» Financial Management and Reporting	» Challenges related to disbursement, financial control, and financial reporting.
			» Auditing	» Challenges caused by weak auditing processes, or excessive auditing procedures.
	Project Data and Monitoring	Delivery challenges caused by ineffective monitoring and evaluation because of 1. Poor data collection and management, 2. Lack of or inappropriate indicators, or 3. Inadequate project supervision.	» Indicators	» Challenges caused by lack of realistic indicators, or duplicating/overlapping indicators, or poorly designed indicators that are misaligned with project objectives.
			» Data Availability and Baselines	» Challenges that stem from a lack of current or accurate data, as well as inability to produce baselines.
			» Reporting and Supervision	» Challenges caused by obstacles in capturing relevant information and reporting it in a timely fashion.

Source: Global Delivery Initiative.

Note: ICT = information and communication technology.

Table A.4. *RAP 2023* Taxonomy of Factors Affecting Project Implementation

Context	Stakeholder	Project
Legislation and regulations	Coordination and engagement	Project design » Objectives (or design) scope » Time allocation or task sequencing » Stakeholder selection » Beneficiary targeting
Governance and politics » Voice and a ccountability » Corruption and patronage » Political interference » Electoral cycles » Rule of law	Commitment and leadership	Project finance » Procurement » Financing mechanism » Budgeting » Financial management and reporting
Conflict and instability	Human resources and organizational capacity » Skilled human resources and organizational capacity » Skill transfer » Staff turnover	Project data and monitoring » Indicators » Data availability and baselines » Reporting and supervision
Social and cultural		Adaptive management
Environment and geography		Risk identification and mitigation
Basic infrastructure » Information and communication technology » Energy and electricity » Transportation		
Disasters and emergency response » Natural disasters » Man-made disasters » Epidemics		
Business environment		

(continued)

Context	Stakeholder	Project
Macroeconomic environment		

Source: Independent Evaluation Group.

Note: RAP = Results and Performance of the World Bank Group.

Data sources. This *RAP* identified implementation factors that were self-reported by project teams in the ICR's section named Factors Affecting Implementation and Performance.

Coding process. The coding team comprised 9 expert evaluators who reviewed about 20 to 40 ICRs each according to the Global Practice of evaluation expertise. The coding process entailed different stages:

1. Training sessions and guiding material. Like in the analysis of outcome types, the *RAP* team held three training sessions for the coders to get familiarized the taxonomy. A coding guideline was used to inform the review of ICR documents and establish clear definitions of key concepts, taking advantage of supplementary materials from DeCODE. The original DeCODE training set was used as an initial frame of reference to guide the coding and to familiarize coders with the operationalization of various categories and subcategories as defined in the DeCODE taxonomy.

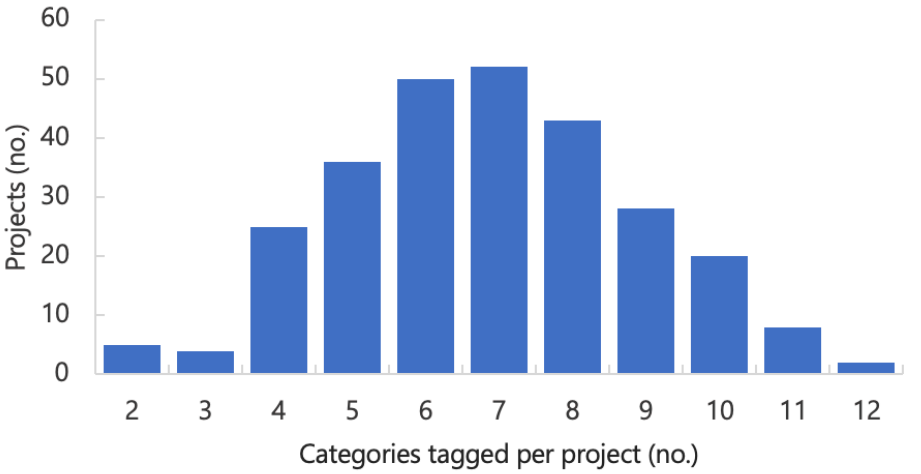
2. Coding software and template. The coding exercise used the NVivo software. The Factors Affecting Implementation and Performance section of the ICR was analyzed by a full read of 273 project reports rather than keyword searches. The coding process involved a sequential approach, where each relevant text extract was classified into categories, followed by subcategories, and ultimately assigned a positive or negative sentiment based on the narrative of the ICR. Individual coders had the opportunity to review and revise their judgments multiple times during this process.

3. Quality assurance. The coding process began with a pilot exercise involving three projects per coder, and pilot outputs were validated to ensure a homogenous understanding of the taxonomy across coders. Several quality assurance steps were incorporated to ensure the accuracy and reliability of the coding output. The *RAP* team collaborated with the methods team

to create intercoder reliability tests, which involved cross-checking coded factors among different team members. At the midpoint of the coding process, the team leader reviewed and provided feedback to coders to ensure the output's consistency, reliability, and validity, and identified and addressed areas for potential recoding or deletion. Progress was tracked using a tool, and final quality assurance tests were conducted after consolidating data from individual coders into the final coding output.

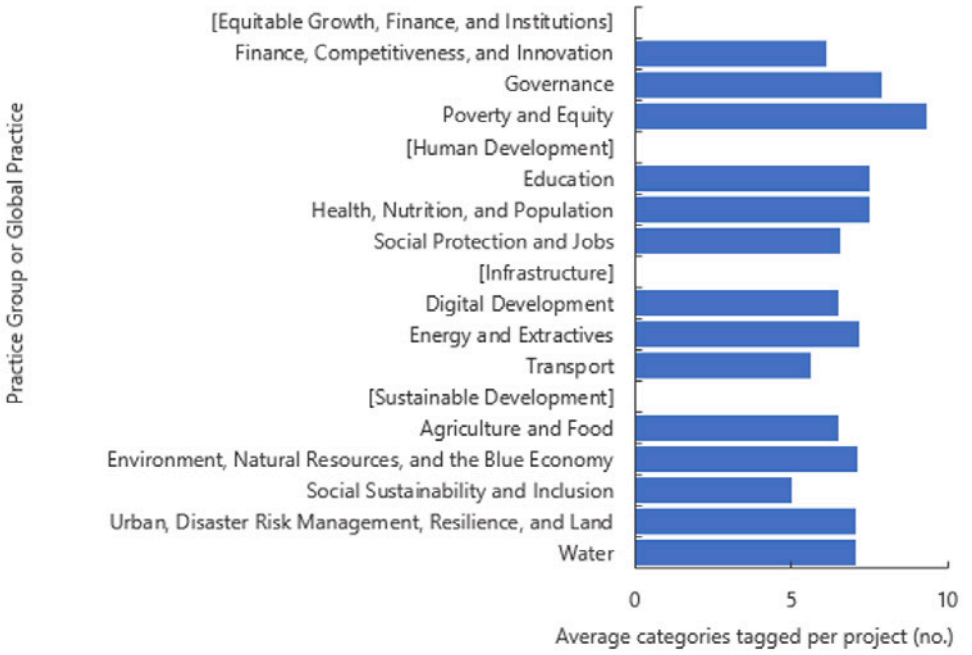
Coding output. The final coding output consisted of 2,479 segments of text (with an average length of 43 words) mapped into relevant categories, subcategories, and sentiment. The resulting data set was well-balanced, having an average of seven tagged factors per project at the category level, which was consistent across all Global Practices and Regions (figures A.6–A.8). This data set served as the training set for the subsequent supervised machine learning exercise, which expanded the analysis of factors affecting implementation to previous years.

Figure A.6. Histogram of Categories Tagged Per Project



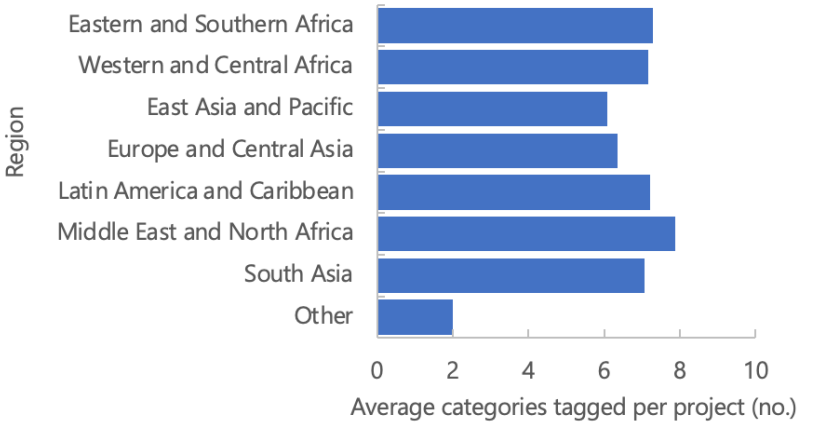
Source: Independent Evaluation Group.

Figure A.7. Categories Tagged per Project, by Practice Group and Global Practice



Source: Independent Evaluation Group.

Figure A.8. Categories Tagged per Project, by Region



Source: Independent Evaluation Group.

Limitations of the analysis of factors affecting implementation. First, factors were identified based on the self-reported narrative of the ICR section concerning factors affecting implementation, which was subject to heterogeneities across different ICR authors. Second, it was not feasible to determine the exact timing of factors within a project's life span based on the information provided in the ICR narratives, except for the subcategory of epidemics, which has a clear starting point in March 2020. The projects reviewed in the FY20–22 cohort were approved between 2003 and 2021, so it cannot be assumed that the pandemic triggered or was directly associated with factors influencing implementation because they may have posed challenges to project implementation before March 2020.

Additional Content Analysis for the Subcategory of Epidemics

Text extractions included in the subcategory of epidemics contained valuable information on how the COVID-19 pandemic specifically affected project implementation, as reported by project teams. We applied an inductive approach to uncover underlying topics embedded in the 443 extracts of text corresponding to 212 projects that reported on the epidemics factor. The content analysis revealed that lockdowns and mobility restrictions, economic downturn, disruptions in services and public institutions operations, implementation delays (caused by supply chain shortages and logistical challenges, and postponement of in-person gatherings, workshops, technical assistance, and capacity-building activities), and reallocation of project funds were the main ways the pandemic affected project implementation (see appendix D for more details).

Supervised Machine Learning

The supervised machine learning exercise used the training set developed in the previous section corresponding to the 273 investment project financing projects included in the *RAP 2023* cohort. To determine whether the implementation factors identified were different from those of previous projects, we used a portfolio of 398 ICRs from the prepandemic period FY18–20 for comparative purposes. At the beginning of the exercise, the prepandemic portfolio

was uncoded, and the team's objective was to predict the corresponding subcategories and sentiments of these projects within the taxonomy.

The analysis comprised the application of two supervised machine learning techniques: (i) a classification model, which was trained by the *RAP* team to learn the taxonomy on a subset of manually coded projects so that the same taxonomy could be applied to a separate portfolio of uncoded projects and (ii) a sentiment analysis model to predict the predominant sentiment (positive or negative) in the text of each of these projects.

Data preprocessing. To prepare the text for further analysis, all text segments included in the training set were converted to lowercase; punctuation signs, leading and trailing white spaces, general and domain-specific "stop words,"[3] and numbers were removed; and all words were lemmatized.[4] Data were also converted to their vector (numerical) representation using term frequency/inverse document frequency technique. [5]

Taxonomy refinement. To ensure that the text extracted for each of the categories and subcategories was distinct enough, the cosine similarity was calculated between each pair of categories and subcategories.[6] The vocabulary for those categories and subcategories with a cosine similarity greater than 50 percent (not distinct enough) was refined by analyzing frequency tables of the unigrams (single words) and bigrams (two subsequent words) included in the vocabulary. Several rounds of vocabulary refinement were completed, including recalculation of cosine similarity after each stage. Two subcategories (skilled human resources and organizational capacity) maintained a high cosine similarity and were therefore merged, resulting in 26 subcategories.

Training set split. The segments of text included in the training set were randomly split into two subsets using an 80:20 ratio: a training set to train classification models and observe their performance, and a testing set to determine how well the chosen model performs outside the model sample.

Classification models. The team applied multiple classification models to the training set (logistic regression, K-nearest neighbors, support vector machine, decision tree, random forest, naïve Bayes, and stochastic gradient descent classifier). Different hyperparameters were tested for each model, and several metrics were calculated to assess model performance.[7,8] The

classification model with the highest accuracy score on the training data was selected (in this case, logistic regression with 75.1 percent accuracy on the training set; table A.5). The best-performing classification model was then applied to the testing set (which had not yet been entered into the models), resulting in an overall accuracy of 75 percent for the testing set. In addition to the overall model accuracy, the team analyzed the accuracy at the subcategory level. Five subcategories (business environment, financial management and reporting, political interference, stakeholder selection, and skill transfer) had an accuracy of 50 percent or less and performed suboptimally in such a way that the model tended to confuse these subcategories with others. To make the model more robust for these subcategories, the team implemented an ensemble approach (one-versus-rest approach) and took the consensus vote.[9]

Model application to unknown data (prepandemic portfolio).

The best-performing classification model was applied to a separate portfolio of 398 projects, the prediction set, for which the team programmatically extracted the text corresponding to the same sections of the ICRs (Key Factors Affecting Implementation and Performance). As noted previously, the team did not know the subcategories or the sentiment corresponding to each of these projects at this stage. Following the same preprocessing steps used for the training set, the text was split into single sentences and analyzed using the classification model. The output of applying the classification model to the prediction set is a probability distribution for each sentence across all 26 subcategories.[10]

Table A.5. Classification Models Applied to the Training Set

Model	Best Hyperparameters	Accuracy (%)
Logistic regression	C=100, penalty=l1, random_state=123, solver=saga	75.13
K-nearest neighbors	n_neighbors: 7, weights: uniform	68.68
Support vector machine	gamma: scale, kernel: linear	73.53
Decision tree	criterion: gini, splitter: random	58.48

(continued)

Model	Best Hyperparameters	Accuracy (%)
Random forest	criterion: gini, n_estimators: 200	71.45
Naïve Bayes	alpha: 0.25, fit_prior: False	72.37
Stochastic gradient descent classifier	alpha: 0.25, fit_prior: False	72.37

Source: Independent Evaluation Group.

Note: The table shows the hyperparameters for each model that performed best and their accuracy on the training set given the best-performing combination of hyperparameters.

Decision criteria for the final assignment of subcategories to each sentence. We defined the following decision criteria:

1. **Sentence length.** Several sentences contained too few words for the model to extract sufficient meaning from that text. The team decided to exclude sentences with fewer than five words.

2. **Highest probability.** Because the prediction of the subcategories was conducted at the sentence level, it was likely that only one subcategory—that with the highest probability for each sentence—truly represented the meaning of the sentence. Therefore, the subcategory with the highest probability was preserved.

3. **Probability threshold.** To address the varying confidence levels of the classification model across different subcategories, as evidenced by the different accuracy scores for each subcategory in the training set, we used the accuracy score for each subcategory in the training set as a threshold to decide which sentences to preserve in the prediction set. This strategy helped the team disregard transition sentences or sentences that addressed topics not included in the taxonomy for which the main predicted subcategory (with the highest probability) had a low probability in absolute terms. In a manual coding setting, a human coder would not have mapped these transition sentences to any subcategory, and the machine learning model consistently acted in a similar way by assigning a low probability to the main predicted subcategory for a given sentence.

4. Exceptions to avoid inclusion and exclusion errors. Two refinements were introduced to the previous rule to avoid errors of inclusion and exclusion. In terms of inclusion errors, a few subcategories had low (50 percent or less) accuracy in the training set (table A.6), and using such a low threshold in the prediction set would result in errors of inclusion. To mitigate this—and taking into consideration the ensemble approach used by the team to make the model more robust for these subcategories—the threshold for these subcategories was set at 51 percent. In terms of exclusion errors, two subcategories achieved very high accuracy in the training set (epidemics, 95 percent; natural disasters, 100 percent). Applying such a high threshold to the prediction set would result in errors of exclusion. To be more conservative, the team reduced the threshold for epidemics to 61 percent and for natural disasters to 86 percent, based on the observed performance of the prediction set.

Table A.6. Accuracy of the Selected Logistic Regression Model for the Training Set for Each Subcategory

Subcategory	Accuracy (%)
Natural disasters	100.00
Epidemics	95.35
Electoral cycles	88.89
Procurement	86.96
Macroeconomic environment	85.71
Risk identification and mitigation	83.33
Commitment and leadership	80.95
Conflict and instability	80.00
Financing mechanism	80.00
Indicators	78.95
Skilled human resources and organizational capacity	77.36
Coordination and engagement	77.08
Beneficiary targeting	75.00
Data availability and baselines	75.00

(continued)

Subcategory	Accuracy (%)
Objectives (or design) scope	74.19
Budgeting	66.67
Legislation and regulations	66.67
Adaptive management	64.00
Time allocation or task sequencing	62.50
Reporting and supervision	60.00
Staff turnover	60.00
Stakeholder selection	50.00
Business environment	44.44
Political interference	33.33
Skill transfer	25.00
Financial management and reporting	22.22

Source: Independent Evaluation Group.

Sentiment analysis. Several pretrained sentiment analysis models (TextBlob, VADER, FLAIR, DistilBERT, and SieBERT) were applied to the training set. Because the team had manually assigned the sentiment to each segment of text included in the training set, this data set was used to select the best-performing model. The best-performing model was SieBERT, with 86.9 percent accuracy (table A.7), so it was applied to each sentence of the prediction set. The output of this step is the assignment of sentiment (positive or negative) to each sentence of the prediction set.

Table A.7. Accuracy of the Training Set for the Sentiment Analysis Models

Model	Accuracy (%)
SieBERT	86.87
FLAIR	84.29
DistilBERT	83.96
VADER	60.50
TextBlob	56.52

Source: Independent Evaluation Group.

Aggregating subcategories at the project level. After identifying sentences relevant for the analysis (by applying the decision rules described previously), results from the classification model and the sentiment analysis model were tabulated. This resulted in a table containing each sentence, the project (P number) from which each sentence was extracted, the subcategory predicted by the classification model, and the sentiment (positive or negative) predicted by the sentiment analysis model. Finally, the project identification number was used to aggregate the results at the project level.

Limitations of supervised machine learning. The classification model had a solid performance with an accuracy of approximately 75 percent and generalized well to unseen text, as evidenced by the very similar performance in the training and testing sets. This also means, however, that the model was likely to misclassify some entries when applied to the prediction set. This is also expected when using a manual approach to coding (due to, for example, fatigue, misunderstanding of labels, or different interpretations of the same codebook). Another limitation is potential biases in the data. To minimize this, the team selected two temporally adjacent cohorts of projects as a prediction set (prepandemic cohort)[11] and a training set (*RAP 2023* cohort) to minimize the linguistic and stylistic differences between both corpora of text. Therefore, the only anticipated difference between the two sets of projects is in connection with the language specific to the COVID-19 pandemic ("Epidemics" category in the taxonomy). In addition, the training set includes references to epidemics other than COVID-19 (such as Ebola, cholera, and measles). Furthermore, the classification model does not perform a keyword search (for words such as "COVID" or "C19") but learns to identify the challenges associated with a pandemic (for example, school closures, travel restrictions, or activity suspensions). This is evidenced by the fact that the model correctly tagged sentences including these examples as "Epidemics" in projects of the prepandemic cohort. Finally, the selected sentiment analysis model SieBERT was a fine-tuned version of the RoBERTa large model, which is a transformers model pretrained on a large corpus of English data. SieBERT was fine-tuned using 15 data sets from diverse sources (such as reviews and tweets). Although this is standard practice in the application of machine learning models—and one that normally correlates with

higher performance—the team completed the additional step of testing all considered sentiment analysis models on the completed training set, which contained 273 ICRs that were manually coded, to ensure that the model translated well to World Bank language. As noted earlier, SieBERT achieved 86.9 percent accuracy when tested on the pandemic data set, which gave the team additional confidence in the capacity of the model to transfer well to a different data set.

Methodology Approach for the International Finance Corporation Analyses

The Independent Evaluation Group's Evaluation Methodology for International Finance Corporation Investment Projects

IEG draws a random stratified representative sample annually from among IFC investment projects approved by the Board of Executive Directors five years earlier that have reached early operating maturity. During the calendar year, IFC investment staff evaluate all active IFC investment projects selected in the sample using Expanded Project Supervision Reports (XPSRs), and IEG independently validates them using Evaluative Notes (EvNotes). For closed projects selected in the sample, IEG prepares a Project Evaluation Summary in lieu of an XPSR. To conduct the project evaluation and validation, IFC and IEG staff refer to XPSR guidelines, which provide the evaluation framework and performance rating criteria.

The evaluation system and performance ratings for IFC investment projects are both objective based and benchmark based. In addition to attention being focused on the achievement of expected objectives stated in the Board report at approval, IFC investment project performance is assessed against several benchmarks, such as performance of peer companies, the market, and similar industries, and considers unintended outcomes (both positive and negative).

The main performance assessment dimensions for IFC investment projects are development outcome, IFC additionality, IFC investment outcome, and IFC work quality. In addition, the XPSR assesses the sustainability of

development and IFC investment outcomes in the longer run by examining project prospects and investment return expectations over the remaining life of the project.

- Development outcome synthesizes a project's performance across four dimensions: project business performance, economic sustainability, environmental and social effects, and private sector development. It is rated on a six-point scale: highly successful, successful, mostly successful, mostly unsuccessful, unsuccessful, and highly unsuccessful.
- IFC additionality assesses the benefit or value addition IFC brings that a client would not otherwise have. It is rated on a four-point scale: excellent, satisfactory, partly unsatisfactory, and unsatisfactory.
- IFC investment outcome assesses the extent to which IFC has realized at the time of evaluation and expects to realize over the remaining life of the investment the loan income, equity returns, or both that were expected at approval. It is rated on a four-point scale: excellent, satisfactory, partly unsatisfactory, and unsatisfactory.
- IFC work quality assesses IFC's operational performance, including in relation to environmental and social aspects, with respect to precommitment work in screening, appraising, and structuring, and its supervision and administration after project approval by the Board and subsequent commitment. It is rated on a four-point scale: excellent, satisfactory, partly unsatisfactory, and unsatisfactory.

The Independent Evaluation Group's Evaluation Methodology for International Finance Corporation Advisory Projects

For all client and sponsor development, and sector development and market creation advisory projects, the IFC advisory services operations staff conduct an evaluation at completion in the form of the Project Completion Report (PCR). IEG validates a random stratified representative sample of these reports each year through Evaluative Notes (EvNotes). IEG annually draws a random stratified representative sample from among projects with PCRs prepared in the previous fiscal year. Both IFC and IEG staff refer to

PCR guidelines when preparing these documents, which provide evaluation frameworks and performance rating criteria. The performance ratings for IFC advisory projects are derived from an objectives-based methodology that establishes minimum thresholds for rating and assessing project effectiveness.

The main performance assessment dimensions for IFC advisory projects are development effectiveness, IFC's role and contribution, and IFC work quality. As part of development effectiveness performance, PCRs assess the sustainability of results over the long term by examining the project's impact achievement beyond the immediate and intermediate outcome achievements.

- Development effectiveness synthesizes a project's performance across five indicators: strategic relevance, output achievement, outcome achievement, impact achievement, and efficiency. It is rated on a six-point scale: highly successful, successful, mostly successful, mostly unsuccessful, unsuccessful, and highly unsuccessful.
- IFC's role and contribution assesses the extent to which IFC added value or made a special contribution to the advisory project. It is rated on a four-point scale: excellent, satisfactory, partly unsatisfactory, and unsatisfactory.
- IFC work quality assesses the extent to which services provided ensured quality at entry and supported effective implementation, through appropriate supervision and execution, toward the achievement of development objectives. IFC work quality and its two dimensions—project preparation and design and project implementation and supervision—are rated on a four-point scale: excellent, satisfactory, partly unsatisfactory, and unsatisfactory.

Analysis, Data Sources, and Sample Coverage

Table A.8 lists data sources and sample coverage of IFC investment and advisory projects used in the *RAP 2023* analyses.

Table A.8. Data Sources and Sample Coverage of IFC Investment and Advisory Projects

Analysis	Data Source	Sample Coverage
IFC project performance ratings	IEG and IFC data	IFC investment projects in CY20–22 XPSR programs and IFC advisory projects in FY20–22 PCR programs evaluated and validated as of June 30, 2023
Analysis of IFC investment project outcome types	IEG and IFC data, XPSR Evaluative Notes	170 IFC investment projects in CY20–22 XPSR evaluation programs evaluated and validated as of December 31, 2022
Factors affecting IFC investment project implementation and performance	IEG data and taxonomy, XPSR Evaluative Notes, IFC data	Same set of projects covered in analysis of outcome types; for prepandemic comparison, *RAP* used 265 investment projects evaluated and validated in CY17–19 XPSR programs

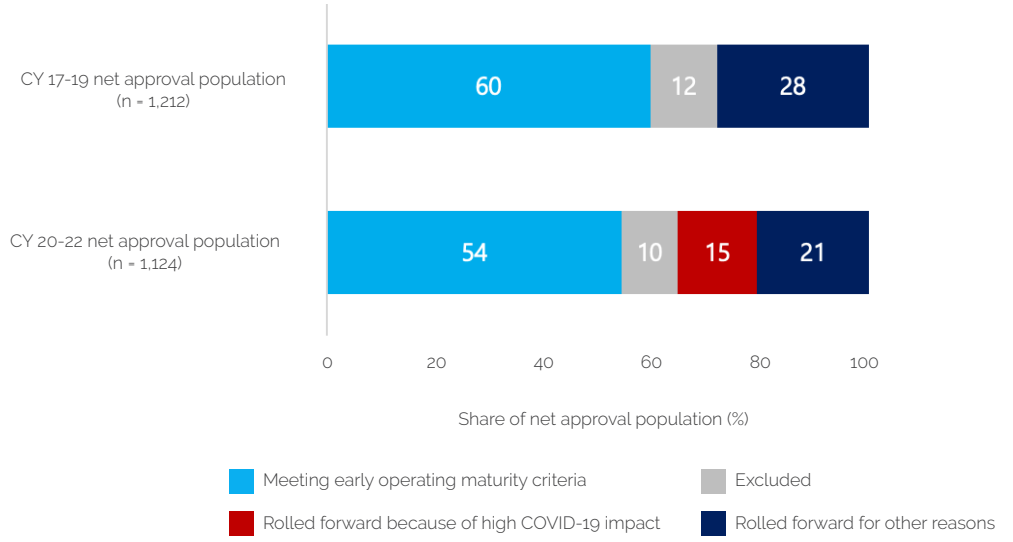
Source: Independent Evaluation Group.

Note: CY = calendar year; FY = fiscal year; IEG = Independent Evaluation Group; IFC = International Finance Corporation; PCR = Project Completion Report; XPSR=Expanded Project Supervision Report.

Sample Selection and Representativeness of International Finance Corporation Investment Projects

The XPSR system is based on a sampling of IFC investment projects that were approved 5 years earlier. The combined sample of calendar year (CY)20–22 XPSR projects was drawn from the net approval population (NAP) of projects approved in CY15–17. At the time of sampling, IEG allowed project evaluations for IFC investment projects that the pandemic highly affected (and some that were moderately affected) to be postponed. Fifteen percent of projects in the population for CY20–22 XPSR cohort were rolled over and were not considered for sampling because COVID-19 highly affected them (figure A.9). Therefore, the projects that COVID-19 highly affected and some others with moderate impact have not yet been evaluated. In addition, 21 percent of projects in this population were rolled over because they were deemed not to be operationally mature. The 36 percent combined share of COVID-19–postponed projects and projects rolled over in the CY20–22 net approval population was higher than the 28 percent of projects rolled over in the prepandemic cohort NAP. Such changes in the sampling processes influenced the profile of projects in the *RAP 2023* cohort to some extent.

Figure A.9. Net Approval Population for CY17–19 and CY20–22 Cohorts According to Share of Projects



Source: Independent Evaluation Group.

Note: CY = calendar year.

IEG selected the stratified random representative sample from investment projects meeting the early operating maturity criteria that had the best fit in terms of representing the population characteristics. In addition to active investment projects, the sample included closed investment projects to represent all mature projects. The overall XPSR sample size was determined to achieve representativeness of the population on a three-year rolling basis, with a sampling error of 5 percent or less at the 95 percent confidence level.

There were in total 232 projects in the combined CY20–22 XPSR programs chosen from a CY15–17 population of 583 projects (sampling rate of 40 percent). A principal goal of sampling is to achieve representativeness, which supports valid performance inferences about the population. Matching of the sample against the population was based on the number of investments. Table A.9 compares the characteristics of the combined sample of 232 CY20–22 XPSRs to those of 583 investment operations in the CY15–17 NAP. There was good fit between the sample and the population and no performance bias. The values of investments shown in table A.9 are for illustrative purposes only.

Table A.9. Representativeness of Combined CY20–22 Expanded Project Supervision Report Sample and CY15–17 Net Approval Population

	Number of Investments					Number of Investments				
	CY20–22 XPSRs (*a*)		CY15–17 NAP (*b*)		(*c*) = (*a*)/(*b*)	CY20–22 XPSRs (*a*)		CY15–17 NAP (*b*)		(c) = (a)/(b)
	No.	%	No.	%	%	Amt.	%	Amt.	%	%
	232	100	583	100	40	9,715	100	23,885	100	41
IFC Commitment										
Mean					-	41.9		41.0		-
Median					-	25.0		25.0		-
Investment Size (US$M)										
Small	49	21	115	20	43	267	3	644	3	42
Medium	146	63	373	64	39	4,631	48	11,736	49	39
Large	37	16	95	16	39	4,817	50	11,505	48	42
	232	**100**	**583**	**100**	**40**	**9,715**	**100**	**23,885**	**100**	**41**
Instruments										
Equity only	63	27	157	27	40	1,575	16	4,161	17	38
Other	169	73	426	73	40	8,140	84	19,724	83	41
	232	**100**	**583**	**100**	**40**	**9,715**	**100**	**23,885**	**100**	**41**

(continued)

	Number of Investments					Number of Investments				
	CY20–22 XPSRs (*a*)		CY15–17 NAP (*b*)		(*c*) = (*a*)/(*b*)	CY20–22 XPSRs (*a*)		CY15–17 NAP (*b*)		(c) = (a)/(b)
Industry Groups										
Financial Institutions Group	88	38	225	39	39	3,683	38	9,564	40	39
Manufacturing, Agribusiness and Services	73	31	183	31	40	2,543	26	6,574	28	39
Infrastructure and Natural Resources	47	20	115	20	41	2,868	30	6,189	26	46
Disruptive Technologies and Funds	24	10	60	10	40	622	6	1,558	7	40
	232	**100**	**583**	**100**	**40**	**9,715**	**100**	**23,885**	**100**	**41**
Regions										
Africa	57	25	135	23	42	1,674	17	3,928	16	43
East Asia and the Pacific	39	17	96	16	41	1,282	13	3,581	15	36
Central Asia and Türkiye	15	6	38	7	39	851	9	2,683	11	32
Europe	23	10	51	9	45	1,020	10	1,930	8	53
Latin America and the Caribbean	54	23	143	25	38	2,105	22	5,659	24	37
Middle East	9	4	24	4	38	850	9	1,258	5	68
South Asia	27	12	70	12	39	1,298	13	3,314	14	39
World	8	3	26	4	31	634	7	1,532	6	41
	232	**100**	**583**	**100**	**40**	**9,715**	**100**	**23,885**	**100**	**41**

(continued)

	Number of Investments					Number of Investments				
	CY20–22 XPSRs (*a*)		CY15–17 NAP (*b*)		(*c*) = (*a*)/(*b*)	CY20–22 XPSRs (*a*)		CY15–17 NAP (*b*)		(c) = (a)/(b)
IDA Status										
IDA and blend	60	26	156	27	38	1,793	18	4,119	17	44
Non-IDA	140	60	351	60	40	6,440	66	16,830	70	38
Global and regional	32	14	76	13	42	1,482	15	2,936	12	50
	232	**100**	**583**	**100**	**40**	**9,715**	**100**	**23,885**	**100**	**41**
Environmental Category										
A	17	7	40	7	43	1,298	13	2,888	12	45
B	92	40	236	40	39	3,776	39	9,326	39	40
C	11	5	25	4	44	130	1	255	1	51
FI	112	48	282	48	40	4,511	46	11,415	48	40
	232	**100**	**583**	**100**	**40**	**9,715**	**100**	**23,885**	**100**	**41**
Indicative Performance										
(i) All investments:[a]										
With loss reserves	10	4	25	4	40	279	3	743	3	38
Without loss reserves	222	96	558	96	40	9,436	97	23,141	97	41
	232	**100**	**583**	**100**	**40**	**9,715**	**100**	**23,885**	**100**	**41**

(continued)

	Number of Investments					Number of Investments				
	CY20–22 XPSRs (*a*)		CY15–17 NAP (*b*)		(*c*) = (*a*)/(*b*)	CY20–22 XPSRs (*a*)		CY15–17 NAP (*b*)		(c) = (a)/(b)
Indicative Performance										
(ii) All investments:[b]										
With write-offs	11	5	24	4	46	455	5	759	3	60
Without write-offs	221	95	559	96	40	9,260	95	23,125	97	40
	232	**100**	**583**	**100**	**40**	**9,715**	**100**	**23,885**	**100**	**41**
Status										
Active	184	79	462	79	40	8,482	87	19,929	83	43
Closed	48	21	121	21	40	1,234	13	3,955	17	31
	232	100	583	100	40	9,715	100	23,885	100	41
Countries (including regional)	75		103							

Source: Independent Evaluation Group.

Note: a. Amounts with loss reserves are the IFC approved investments that are affected by loss reserves (and not the actual amount reserved).
b. Amounts with write-offs are the IFC approved investments that are affected by write-offs (and not the actual amount written-off).
CY = calendar year; IDA = International Development Association; NAP = net approval population; XPSR = Expanded Project Supervision Report.

Table A.10. Inferences Drawn from the CY20–22 Expanded Project Supervision Report Sample about the CY15–17 Net Approval Population

Indicator	Sample Success Rate	Estimate of Population Success Rate[a]	Standard Error[b]	Sample Error[c]	95% Confidence Interval	
					Lower bound	Upper bound
Development outcome	0.50	0.51	0.02	0.05	0.46	0.56
1. Project business performance	0.50	0.51	0.03	0.05	0.46	0.56
2. Economic sustainability	0.44	0.45	0.03	0.05	0.40	0.50
3. Environmental and social effects	0.68	0.68	0.03	0.06	0.62	0.73
4. Private sector development	0.60	0.60	0.03	0.05	0.55	0.65
IFC investment outcome	0.60	0.61	0.02	0.05	0.56	0.66
5. Loan	0.76	0.67	0.03	0.05	0.62	0.73
6. Equity	0.34	0.31	0.03	0.06	0.25	0.36
IFC work quality	0.55	0.55	0.03	0.05	0.50	0.60

(continued)

Indicator	Sample Success Rate	Estimate of Population Success Rate[a]	Standard Error[b]	95% Confidence Interval		
				Sample Error[c]	Lower bound	Upper bound
7. Screening, appraisal, and structuring	0.54	0.54	0.03	0.05	0.49	0.59
8. Supervision and administration	0.67	0.68	0.02	0.05	0.64	0.73
IFC additionality						
9. Additionality	0.54	0.55	0.03	0.05	0.50	0.60

Source: Independent Evaluation Group.

Note: CY = calendar year; IFC = International Finance Corporation.
a. Estimates of the success rate of operations in the underlying population are based on the actual success rate of operations in the sample and uses equation 5.8 in Levy and Lemeshow (1991, 112).
b. Standard error of the estimated success rate of operations in the population, shown in column 4, is based on equation 6.8 in Levy and Lemeshow (1991, 121).
c. Sampling error is computed as 1.96 x standard error (1.96 is the t-statistic associated with the 95% confidence level).

Table A.10 presents the results for the nine rated indicators and the four dimensions for the CY15–17 NAP based on the CY20–22 XPSR samples. At ±5 percent to 6 percent, the sampling error for some estimates is slightly higher than the ±5 percent range specified in the relevant Multilateral Development Bank Evaluation Cooperation Group good practice standard. This is because not all CY22 XPSRs have been evaluated, which led to a marginally higher sampling error than the standard. Even with the limitation, the CY20–22 sample success rates can be mostly attributed to those of the CY15–17 NAP within the limits shown in last two columns of table A.10.

Sample Selection and Representativeness of International Finance Corporation Advisory Projects

At implementation completion, IFC prepares the PCR for all client and sponsor development, and sector development and market creation advisory projects. Each year, IEG validates a random, stratified, representative sample of projects with PCRs prepared in the previous fiscal year. The coverage rate is determined to be sufficient to allow for statistical inference about (development effectiveness) success rates in the population and achieve representativeness on a three-year rolling basis with a sampling error of 5 percent or less at the 95 percent confidence level. The stratified random sample has the best fit in terms of representing the population characteristics.

There were 185 PCRs in the combined FY20–22 samples, chosen from a population of 353 projects (sampling rate of 52 percent). As with XPSR sampling, the principal goal of PCR sampling has been representativeness to support valid performance inferences about the population. Table A.11 compares the characteristics of the combined sample of 185 FY20–22 PCRs with those of 353 advisory operations in the FY20–22 NAP. Overall, there was generally close alignment of characteristics between the sample and the NAP.

Table A.11. Representativeness of Combined FY20–22 Project Completion Report Sample versus Net Approval Population

PCR Sample vs PCR NAP	Number of Advisory Projects					Value of Funds Managed by IFC (actual US$)				
	FY20–22 PCRs (a)		*PCR NAP (b)*		(c) = (a)/(b)	*FY20–22 PCRs (a)*		*PCR NAP (b)*		(c) = (a)/(b)
	No.	%	No.	%	%	Amt.	%	Amt.	%	%
	185	100	353	100	52	US$339,516,828	100	US$649,653,639	100	52
Total Funds Managed by IFC										
Mean						1,835,226		1,840,379		
Median						1,319,606		1,284,451		
Funding Size (actual US$)										
Small	41	22	78	22	53	13,596,722	4	26,684,060	4	51
Medium	110	59	211	60	52	170,887,342	50	322,361,003	50	53
Large	34	18	64	18	53	155,032,764	46	300,608,576	46	52
	185	**100**	**353**	**100**	**52**	**339,516,828**	**100**	**649,653,639**	**100**	**52**
Project Duration (months)										
Short	41	22	83	24	49	30,935,460	9	64,326,120	10	48
Medium	107	58	199	56	54	194,715,038	57	367,205,188	57	53
Project Duration (months)										
Long	37	20	71	20	52	113,866,330	34	218,122,331	34	52

(continued)

PCR Sample vs PCR NAP	Number of Advisory Projects					Value of Funds Managed by IFC (actual US$)				
	FY20–22 PCRs (a)		*PCR NAP (b)*		(c) = (a)/(b)	*FY20–22 PCRs (a)*		*PCR NAP (b)*		(c) = (a)/(b)
	No.	%	No.	%	%	Amt.	%	Amt.	%	%
	185	100	353	100	52	US$339,516,828	100	US$649,653,639	100	52
	185	**100**	**353**	**100**	**52**	**339,516,828**	**100**	**649,653,639**	**100**	**52**
Country Borrower Type										
IDA and blend	106	57	204	58	52	179,421,228	53	347,432,383	53	52
Non-IDA	60	32	112	32	54	117,556,379	35	201,607,253	31	58
Global and regional	19	10	37	10	51	42,539,221	13	100,614,003	15	42
	185	**100**	**353**	**100**	**52**	**339,516,828**	**100**	**649,653,639**	**100**	**52**
Country FCS Status										
FCS	31	17	60	17	52	55,991,752	16	104,266,950	16	54
Non-FCS	154	83	293	83	53	283,525,076	84	545,386,689	84	52
	185	**100**	**353**	**100**	**52**	**339,516,828**	**100**	**649,653,639**	**100**	**52**
Project Type										
Client and sponsor development	64	35	121	34	53	83,885,473	25	139,986,774	22	60
Sector development and market creation	121	65	232	66	52	255,631,355	75	509,666,865	78	50
	185	**100**	**353**	**100**	**52**	**339,516,828**	**100**	**649,653,639**	**100**	**52**

(continued)

PCR Sample vs PCR NAP	Number of Advisory Projects					Value of Funds Managed by IFC (actual US$)				
	FY20–22 PCRs (a)		*PCR NAP (b)*		*(c) = (a)/(b)*	*FY20–22 PCRs (a)*		*PCR NAP (b)*		*(c) = (a)/(b)*
	No.	%	No.	%	%	Amt.	%	Amt.	%	%
	185	100	353	100	52	US$339,516,828	100	US$649,653,639	100	52
Primary Business Areas										
Financial Institutions Group	63	34	123	35	51	106,905,095	31	189,859,248	29	56
Equitable Growth, Finance and Institutions	29	16	61	17	48	46,779,407	14	122,771,155	19	38
Public-Private Partnership	19	10	37	10	51	35,153,847	10	57,556,334	9	61
Environment Social and Governance	19	10	34	10	56	38,119,881	11	69,510,253	11	55
Regional Advisory	18	10	35	10	51	39,818,074	12	84,781,506	13	47
Manufacturing, Agribusiness and Services	13	7	24	7	54	31,439,149	9	48,869,876	8	64
Infrastructure and Natural Resources	9	5	15	4	60	15,171,219	4	34,399,759	5	44
Disruptive Technologies and Funds	2	1	3	1	67	1,000,000	0.3	2,621,804	0.4	38
Others	13	7	21	6	62	25,130,156	7	39,283,704	6	64
	185	**100**	**353**	**100**	**52**	**339,516,828**	**100**	**649,653,639**	**100**	**52**

(continued)

PCR Sample vs PCR NAP	Number of Advisory Projects					Value of Funds Managed by IFC (actual US$)				
	FY20–22 PCRs (a)		*PCR NAP (b)*		(c) = (a)/(b)	*FY20–22 PCRs (a)*		*PCR NAP (b)*		(c) = (a)/(b)
	No.	%	No.	%	%	Amt.	%	Amt.	%	%
	185	100	353	100	52	US$339,516,828	100	US$649,653,639	100	52
Regions										
Africa	66	36	126	36	52	102,026,359	30	193,236,979	30	53
Central Asia and Türkiye	7	4	15	4	47	18,068,291	5	37,910,319	6	48
East Asia and the Pacific	35	19	67	19	52	84,847,030	25	148,617,436	23	57
Europe	12	6	24	7	50	39,916,222	12	69,303,381	11	58
Latin America and the Caribbean	19	10	38	11	50	21,642,360	6	51,723,689	8	42
Middle East	16	9	28	8	57	22,991,669	7	36,849,653	6	62
South Asia	24	13	44	12	55	40,741,342	12	80,236,546	12	51
World	6	3	11	3	55	9,283,555	3	31,775,636	5	29
	185	**100**	**353**	**100**	**52**	**339,516,828**	**100**	**649,653,639**	**100**	**52**
Countries (including regional)	**77**		**95**							

Source: Independent Evaluation Group.

Note: FY = fiscal year; FCS = fragile and conflict-affected situation; IDA = International Development Association; NAP = net approval population; PCR = project completion report.

Table A.12. Inferences Drawn from the FY20–22 Project Completion Report Sample about the Net Approval Population

Indicator	Sample Success Rate	Estimate of Population Success Rate[a]	Standard Error[b]	Sample Error[c]	95% Confidence Interval	
					Lower bound	Upper bound
A. Development effectiveness	0.54	0.54	0.03	0.05	0.49	0.59
1. Strategic relevance	0.70	0.71	0.02	0.05	0.66	0.76
2. Output achievement	0.86	0.87	0.02	0.03	0.83	0.90
3. Outcome achievement	0.54	0.54	0.03	0.05	0.49	0.59
4. Impact achievement	0.23	0.24	0.03	0.06	0.18	0.30
5. Efficiency	0.54	0.54	0.03	0.05	0.49	0.60
B. IFC's role and contribution	0.82	0.82	0.02	0.04	0.78	0.86
C. IFC overall work quality	0.59	0.59	0.03	0.05	0.54	0.64
6. Project preparation and design	0.49	0.49	0.03	0.05	0.44	0.54

(continued)

Indicator	Sample Success Rate	Estimate of Population Success Rate[a]	Standard Error[b]	Sample Error[c]	95% Confidence Interval	
					Lower bound	Upper bound
7. Project implementation and supervision	0.61	0.61	0.03	0.05	0.56	0.66

Source: Independent Evaluation Group.

Note: IFC = International Finance Corporation.
a. Estimates of the success rate of operations in the underlying population are based on the actual success rate of operations in the sample and uses equation 5.8 in Levy and Lemeshow (1991, 112).
b. Standard error of the estimated success rate of operations in the population, shown in column 4, is based on equation 6.8 in Levy and Lemeshow (1991, 121).
c. Sampling error is computed as 1.96 × standard error (1.96 is the t-statistic associated with the 95 percent confidence level).

Table A.12 presents the results for the seven rated indicators and the three dimensions for the FY20–22 NAP estimated using the FY20–22 PCR samples. The sampling error for some estimates is slightly greater than ±5 percent, the range specified in the relevant Multilateral Development Bank Evaluation Cooperation Group good practice standard. This is because not all FY22 PCRs have been validated, which leads to a higher sampling error than the standard. Even with the higher sampling error, the FY20–22 sample success rates can be attributed to those of the NAP within the limits shown in last two columns of table A.12.

Factors Affecting Implementation and Performance

To identify the factors influencing IFC investment project implementation and performance during the COVID-19 pandemic, the *RAP 2023* team performed a qualitative review and content analysis of project evaluation documents. This included 170 IFC investment projects in CY20–22 for which the evaluation and validation were completed by the cutoff date of December 31, 2022. For each project, the *RAP* team identified the top three factors that positively or negatively affected project performance and classified them using the existing taxonomy of performance factors, consisting of 5 categories and 51 subcategories developed by IEG (table A.13). The taxonomy was based on common challenges and issues faced in more than 1,000 evaluated IFC investment projects. For these projects, IEG had used machine learning in addition to human thinking to identify key performance factors and classify categories and subcategories. This machine learning model was fully tested for IFC's Financial Institutions Group investment projects and partially tested for IFC's Infrastructure and Natural Resources industry group investment projects to ensure a reasonable level of accuracy of predictions made by the machine learning model. This *RAP* contributed to further training for the machine learning model with the data collected by humans for the CY20–22 *RAP* cohort. The current accuracy rate of the machine learning model is 71 percent, which means that the model is identifying the same top performance factors as humans in 71 percent of projects.

For CY20–22 projects, the factor identification and classification exercise involved two steps. First, the *RAP 2023* team conducted its factor analysis by

identifying the top three factors for each project based on review of project evaluation documents and classified them according to the taxonomy. Second, for each reviewed project, the Financial and Private Sector Micro Unit sector leaders reviewed and validated these key factors and their categories and subcategories. The second step ensured not only appropriate classification of categories and subcategories but also correct identification of factors that contributed to project performance. An additional review across industries made sure that classifications were consistent over the total portfolio of EvNotes analyzed. While this taxonomy was implemented, the team introduced two additional subcategories, on which the model had not been previously tested extensively, to better reflect factors associated with IFC's investment funds projects.

For prepandemic projects in the CY17–19 cohort, the *RAP* used supervised-learning algorithms to identify key factors affecting performance. Based on the results of testing different algorithms (naïve Bayes, random forest, support vector machine, and multilayer neural network), we decided to use naïve Bayes for categorizing paragraphs from the project evaluation documents and, specifically, for assigning a probability that a particular paragraph would be assigned to a particular category in the taxonomy. To allow paragraphs to be categorized in more than one theme, the classification assigned a primary, secondary, and tertiary subcategory in addition to a probability of assignment to each. As an additional measure to aid categorization, we also used a sentiment analysis to assign a score to each paragraph ranging from –1 (totally negative; paragraph includes information on a factor or issue that is a barrier or impediment to project implementation) to +1 (totally positive; paragraph includes information on a factor or aspect that contributes to success in project implementation). This analysis was conducted using polarity scores from Python's Natural Language Processing Package. Streamlining and refining model subcategories involved additional diagnostics such as cosine similarity. In the case of high similarity scores, we checked keywords and categories to ensure that the groups identified in the taxonomy were (to the extent possible) mutually exclusively defined. After a few iterations, we were able to eliminate several categories with problematic overlaps, further improving the subcategories in the taxonomy (for details on the machine learning application, refer to Bravo et al. 2023).

Table A.13. Taxonomy of Performance Factors (Categories and Subcategories)

No.	Category	Subcategory	Definition
1	Country and macro factors	Civil unrest and armed conflict	Factors related to civil unrest, armed conflict, and war
2		Economic factors	Factors related to the macroeconomic environment, inflation, monetary policy, or austerity measures
3		Epidemics and COVID-19	Factors related to epidemics (human, animal and plant) and COVID-19
4		Expropriation, Nationalization, and Transferability	Factors related to expropriation, nationalization, transfer, and convertibility
5		Foreign exchange and local currency factors	Factors related to currency fluctuation, exchange rate and local currency issuance instruments
6		Legal or regulatory factors	Factors related to regulatory policies, government, legislation, and bureaucratic mechanisms
7		Natural disasters	Factors related to natural disasters such as hurricanes and earthquakes
8		Political factors	Factors related to the political environment, including legislative and electoral dynamics
9	Market, sector, and industry factors	Business factors	Factors related to business model, cyclical business, or the operating environment
10		Competition	Factors related to market competition: barriers to entry, monopolies, market dominance, and penetration
11		Customers	Factors related to identifying correct target markets and clientele
12		Market share	Factors related to market share

(continued)

No.	Category	Subcategory	Definition
13		Pricing	Factors related to price elasticity, supply, and marginal gains
14	Sponsor or client (management, sponsorship, and leadership)	Capacity, capitalization, leverage	Factors related to sponsor capacity, capitalization, and leverage
15		Commitment and motivation	Factors related to the strength and valence of strategic alignment, including issues of compatibility, motivation, and ownership
16		Conflicts of interest, corporate governance	Factors related to minority interest, conflicts of interest and corporate governance
17		Integrity, transparency, fairness, reputation	Factors related to integrity and transparency, such as disclosures of sensitive ethical issues, irregularities, and negative public perceptions
18		Organizational structure	Factors related to organizational culture, institutional procedures, policies, and accountability
19		Technical expertise, track record, and capacity	Factors related to the quality and expertise of the management team, their technical skills and track record, and contractor competency, familiarity, and acumen
20		Succession	Factors related to succession, especially in family-owned businesses
21	Project-inherent challenges	Asset quality	Factors related to asset quality
22		Cost overruns and delays	Factors related to overruns or delays
23		Earnings and profitability	Factors related to earnings and profitability

(continued)

No.	Category	Subcategory	Definition
24		Environment and sustainability	Factors related to environmental standards, social health and safety parameters, or other safety standards
25		Expansion	Factors related to acquisition, modernization, and expansion
26		Funding	Factors related to funding
27		Greenfield	Factors related to greenfield projects
28		Gender	Factors related to gender
29		Liquidity	Factors related to liquidity
30		Technology	Factors related to changes in technology that impact project performance
31		Training, know-how, and implementation	Factors related to training and know-how
32	Other	Additionality principle and catalytic Role	Factors related to additionality and added value
33		Coordination and collaboration with World Bank, other DFIs, donors, and other external stakeholders	Factors related to combined partnership and collaboration among the various stakeholders: the World Bank, donors, DFIs, and other external stakeholders
34		Coordination and collaboration within IFC: AS-IS	Factors related to use of investment and advisory services to enhance IFC roles and contributions
35		Project scoping and screening, country and stakeholder assessment, client needs assessment	Factors related to ex ante market analysis, due diligence, and consumer preferences
36		Client selection, commitment, and capacity	Factors related to client or implementing-partner selection (appropriateness) and client commitment and involvement

(continued)

No.	Category	Subcategory	Definition
37		Project design	Factors related to project design
38		Financial model, project cost, and sensitivity assumptions	Factors related to financial modeling assumptions, including issues regarding overambitious objectives, deviations from forecasting estimates, and scaling
39		Market assessment	Factors related to market assessment, market analysis, and consumer preferences
40		Resources and timeline	Factors related to staffing, budget, and timeline
41		Supervision and reporting	Factors related to supervision and reporting; and (ii) taking measures to enhance these, as well as proactive client and stakeholder follow-up
42		Sensitivity analysis	Factors related to sensitivity analysis, worst case scenarios, stress tests, risks to achieving development outcomes
43		Documentation	Factors related to the quality of monitoring, documentation, and reporting
44		Loan issues	Factors related to loan agreements, operating policies, breaches, or technical defaults
45		Relationship management	Factors related to the quality and scope of relationship management, including fruitful and proactive engagements with on-site staff
46		Debt issues	Factors related to debt issues, such as syndication, repayment, security, and refinancing
47		Equity issues	actors related to equity, valuation, and shareholder rights

(continued)

No.	Category	Subcategory	Definition
48		Financial risk mitigation	Factors related to risk-mitigation mechanisms such as guarantees, securities, prepayment penalties, and restructuring mechanisms
49		Prepayments	Factors related to prepayments
50		Monitoring and evaluation	Factors related to compliance, monitoring, including measurement, reporting, auditing, monitoring and evaluation plan and framework, appropriate indicators and targets, and clarity of data collection and evaluation approach
51		Other issues	Factors related to other issues

Source: Independent Evaluation Group.

Note: AS = advisory services; DFI = development finance institution; IFC = International Finance Corporation; IS = investment services.

Using the collected data, the *RAP* team analyzed the prevalence of key factors that contributed to some projects performing better or worse than others in the context of the COVID-19 pandemic. The team also explored whether there were fundamental changes in the pandemic context by comparing factors in CY20–22 projects exposed to COVID-19 with factors in CY17–19 projects not exposed to COVID-19 at the time of evaluation. The team analyzed the similarities and differences of the main factors across projects or country groups.

Outcome Types

Defining Outcome Types for International Finance Corporation Investment Projects

RAP 2021 developed a 13-category typology of intended outcomes that leveraged IFC's Anticipated Impact Measurement and Monitoring (AIMM) system. These intended outcomes were aligned with those defined by the AIMM sector frameworks. AIMM sector frameworks have been developed for

more than 20 key sectors and subsectors (for example, small and medium enterprise finance, manufacturing, power, and private equity funds) of IFC's investment operations. The AIMM system identifies key development outcomes (defined as outcome claims) with specific indicators for each investment project, in accordance with the theory of change defined in each AIMM sector framework. Each sector framework identifies an expected theory of change that indicates how the projects in each relevant sector are expected to address development gaps. This is done by demonstrating typical outcomes to be achieved by each project at both the project and market level. Each sector framework also includes a list of standard indicators and categorizes them under specific outcome types. Based on the impact thesis and list of indicators, *RAP 2021* developed an outcome typology for 13 outcome categories and some subcategories. *RAP 2021* added additional categories that were not specified in the AIMM sector frameworks to compile 28 outcome types. *RAP 2021* identified outcome claims for projects based on their backfilled AIMM worksheet, doing this for all the IFC investment projects that were self-evaluated by IFC and validated by IEG between CY12 and CY19.

RAP 2023 leveraged the outcome typology developed by *RAP 2021* and applied it to all IFC projects that were self-evaluated by IFC and validated by IEG between CY20 and CY22, but only those projects with XPSRs validated by December 2022 were included in the analysis. Because no AIMM worksheets (where project outcome claims were included) were available for most projects in this cohort, *RAP 2023* reviewed the text of IEG EvNotes and coded descriptions of project- and market-level development outcomes that the projects were intended to achieve.

The *RAP* team reviewed the AIMM sector frameworks, which has remained the same since *RAP 2021*, suggesting that the outcome typology developed in *RAP 2021* was still relevant. However, *RAP 2023* enhanced *RAP 2021's* outcome typology by adding new subcategories and revising definitions of some subcategories. *RAP 2023* identified 33 outcome types (28 project level and 5 market level; table A.14). *RAP 2023* included only outcome claims that were clearly identified in the EvNote to capture key objectives based on what the IEG evaluator had already judged were the main intended objectives. A small number of outcome claims were not accompanied by specific indicators to

measure their results. IEG shared its outcome analysis approach with IFC in the Concept Note and responded to IFC's questions about the process.

Table A.14. Outcome Typologies for IFC Investment Projects

Outcome Type	Description
1.1 - Access to goods and services	Increase in number of final beneficiaries of goods and services of the project or company; increase in volume of goods and services produced by project or company
1.1.1 - Access to goods and services (MSME)	Increase in number of MSMEs as final beneficiaries of goods and services of the project or company; increase in volume of goods and services produced or provided by project or company
1.1.2 - Access to goods and services (gender)	Increase in number of final female beneficiaries of goods and services of the project or company; increase in volume of goods and services produced or provided by project or company
1.1.3 - Access to goods and services (customers)	Increase in number of individual customers as final beneficiaries of goods and services of the project or company; increase in volume of goods and services produced or provided by project or company
1.1.4 - Access to goods and services (miscellaneous)	Increase in number of final beneficiaries of goods and services of the project or company other than MSMEs, female beneficiaries, and individual customers or a mix of these final beneficiaries; increase in volume of goods and services produced by project or company
1.1.5 - Access to goods and services (direct client level)	Increase in capacity of project or direct client company to produce goods and services because of IFC investment
1.2 - Quality and affordability of goods and services	Improved quality of goods and services produced by project or company compared with baseline or with other producers or providers. Lower production costs or process. Reduced prices of goods and services compared with the baseline or other producers or providers
1.2.1 - Quality of goods and services	Improved quality of goods and services produced by project or company compared with the baseline or other producers or providers
1.2.2 - Affordability of goods and services	Reduced prices of goods and services compared with the baseline or other producers or providers
1.2.3 - Increased efficiency of direct client company	Lower production costs or processes of project or company

(continued)

Outcome Type	Description
1.3 - Increased capacity of final beneficiaries	Enhanced capacity of final beneficiaries as a result of advisory services or training that is part of project scope
1.4 - Improved living standards (earnings) of individuals	Increase in revenue or decrease in expenditures by final beneficiaries (individuals) of goods and services produced by the project or company
1.5 - Improved sales or profitability of enterprises	Increase in revenue, decrease in expenditures, or increase in overall productivity by final beneficiaries (enterprises) of goods and services produced by project or company
2.1 - Suppliers and distributors reached	Increase in number of suppliers who provide inputs to project or company or expansion of network of distributors of goods or services produced by project or company
2.2 - Improved capacity of suppliers and distributors	Increase in capacity of suppliers or distributors as a result of advisory services or training that is part of project scope
2.3 - Improved sales and profitability of suppliers and distributors	Increase in volume of inputs provided by suppliers or increase in the goods or services to be distributed by its distributors
3.1 - Increased employment	Increase in direct employment of client company
3.2 - Improved capacity and skills	Training provided to employees of project or company
3.3 - Improved earnings of employees	Increase in wages to employees of project or company
4.1 - Increased transfers to government	Increase in payments by project or company to government, such as in the form of taxes, royalties, fees, or dividends
5.1 - Increased money spent or transferred to community	Increase in payments to communities around the project or company, such as on health, educational, or vocational programs
6.1 - Enhanced environmental and social standards of the client	Improvement in environmental and social standards by IFC
6.2 - Greenhouse gas emissions	Decrease in or avoidance of greenhouse gas emissions

(continued)

Outcome Type	Description
6.3 - Efficient use of resources	Decrease in use of water and other resources, improvement in solid waste management, implementation of waste-to-energy project
7.1 - Gross value added	Gross value added to economy (calculated based on a multiplier and expressed in monetary value)
7.2 - Induced or indirect employment	Induced and indirect employment based on multipliers
7.3 - Exports	Increase in exports of goods and services, generating foreign currency
8.1 - Governance	Improvement in corporate governance or increase in capacity of client company
9 - Competition in the market	Increase in ability of firms to enter, exit, compete, innovate, and strive for efficiency under fair and good regulatory governance; price changes; new practices, technology, product innovation (first movers); product and business model differentiation, change in product offering, value addition; increase in efficiency under fair and good regulatory governance
10 - Resilience in the market	Increase in market depth and improvement in market structure, regulation, and governance to help markets withstand physical, financial, economic, or climate-related shocks; improved corporate governance of direct clients; diversification (for example, energy sources or funding sources in sectors or products); increase in capacity to face shocks and stress; increase in market depth and improvement in market structure, regulation, and governance (capacity of regulator); decrease in domestic supply volatility; increase in energy security; increase in financial stability and consumer protection
11 - Integration in the market	Increase in physical or financial connectivity to support greater market integration, greater integration with financial markets and domestic and global value chains, enhanced physical or financial connectivity, geographical integration, integration with financial markets (including capital mobilization), data integration, growing domestic and global value chains, trade diversification, economic complexity
12 - Inclusiveness in the market	Increase in fair and full access to all goods, services, finance, and economic opportunities, including for underserved groups; increased inclusiveness and improved access; establishment of market-wide enabling framework or standards supporting inclusive business; increase in diversity

(continued)

Outcome Type	Description
13 - Sustainability in the market	Adoption of climate-related, environmentally and socially sustainable products, technologies, and practices; increased ability of firms and industries to apply environmentally and socially sustainable approaches to mitigate risk, realize opportunities, and maximize operational efficiency; adoption of climate-related, environmentally and socially sustainable products, technologies, standards, and practices; development of legal or regulatory framework that fosters sustainability; broad capacity and supporting institutions or sustainability practice

Source: Independent Evaluation Group.

Note: IFC = International Finance Corporation; MSME = micro, small, and medium enterprise.

Methodology for Analysis on Outcomes Achieved

Because AIMM had assessed only a small share of projects in the *RAP 2023* cohort, *RAP 2023* assessed the extent to which expected outcomes were achieved at evaluation by verifying the results presented in the project EvNote. An outcome was considered fully achieved, partially achieved, not achieved, or cannot be verified based solely on the text of the project EvNote, which itself validated the project's self-evaluation XPSR. The *RAP 2023* team did not apply any additional judgment, assessment, or methodology.

Methodology Approach for the Multilateral Investment Guarantee Agency Analyses

For each MIGA guarantee project that has reached early operating maturity, MIGA underwriting staff conduct a self-evaluation by preparing a Project Evaluation Report (PER) that IEG independently validates through a PER Validation Note (ValNote). To conduct the project evaluation and validation, both MIGA and IEG staff refer to PER guidelines, which provide the evaluation framework and performance rating criteria. The evaluation system and performance ratings for MIGA projects are both objectives and benchmarks based. In addition to attention being focused on the achievement of expected objectives stated in the president's report at approval, the performance of MIGA guarantee projects is assessed against several benchmarks (such as performance of peer companies, the market, and similar industries) and considers unintended outcomes (both positive and negative).

The main performance assessment dimensions for MIGA guarantee projects are development outcome, MIGA's role and contribution, and MIGA work quality. The PER also assesses the sustainability of development outcomes in the long term by examining the project's prospects over its remaining life.

- Development outcome measures performance across four indicators: project business performance, economic sustainability, environmental and social effects, and foreign investment effects. It is rated on a six-point scale: highly successful, successful, mostly successful, mostly unsuccessful, unsuccessful, and highly unsuccessful. Up until FY19, the ratings were based on a four-point scale: excellent, satisfactory, partly unsatisfactory, and unsatisfactory.
- MIGA's effectiveness synthesizes MIGA's performance across three indicators: project strategic relevance, MIGA's role and contribution, and MIGA work quality. It is rated on a four-point scale: excellent, satisfactory, partly unsatisfactory, and unsatisfactory.
- MIGA's role and contribution assesses the benefits and value-added that MIGA brings to the client, the project, or the political risk insurance industry. It is rated on a four-point scale: excellent, satisfactory, partly unsatisfactory, and unsatisfactory.
- MIGA work quality addresses due diligence and underwriting processes, including of risk assessment and mitigation, and monitoring after the issuance of the MIGA guarantee. It is rated on a four-point scale: excellent, satisfactory, partly unsatisfactory, and unsatisfactory.

Analysis, Data Sources, and Sample Coverage

Table A.15 indicates the data sources and sample coverage of MIGA guarantee projects used in the *RAP 2023* analyses.

Table A.15. Data Sources and Sample Coverage of MIGA Guarantee Projects

Analysis	Data Sources	Sample Coverage
MIGA guarantee project performance ratings	IEG and MIGA data	MIGA guarantee projects in FY17–22 PER programs evaluated and validated as of June 30, 2023
Analysis of MIGA guarantee project outcome types	IEG and MIGA data	16 MIGA guarantee projects in FY20–22 PER programs evaluated and validated as of December 31, 2022
Factors affecting MIGA guarantee project implementation and performance	IEG and MIGA data, PER Validation Notes	Same set of guarantee projects covered in analysis of outcome types

Source: Independent Evaluation Group.

Note: IEG = Independent Evaluation Group; MIGA = Multilateral Investment Guarantee Agency; PER = Project Evaluation Report.

During the pandemic, delivery of MIGA self-evaluations was delayed. Self-evaluations for seven projects in the FY22 program were received by IEG only recently in the third quarter of FY23. Therefore, validations by IEG have not yet been completed for these projects. Delays in delivery of self-evaluations were related to client reporting delays and challenges in visiting project sites, which constrained information gathering and meant it took more time to fill information gaps. In addition, MIGA self-evaluations for one project in FY21 and nine projects in FY22 have been deferred to the FY23 evaluation program. Such delays in the delivery of self-evaluations influenced the profile of projects analyzed in the *RAP 2023* cohort to some extent.

Factors Affecting Implementation and Performance

To identify the factors influencing MIGA guarantee project implementation and performance during the COVID-19 pandemic, the *RAP 2023* team performed a qualitative review and content analysis of project evaluation documents. This included 16 MIGA guarantee projects, for which the evaluation and validation was completed by the cutoff date of December 31, 2022. For each project, the *RAP* team identified the top three factors that positively or negatively affected project performance and classified them using the

existing taxonomy of performance factors, consisting of 5 categories and 51 subcategories developed by IEG. For the performance factor analysis of MIGA guarantee projects, the same methodology was used for the *RAP* as for IFC investment projects (for taxonomy and other details, see the Methodology Approach for the International Finance Corporation Analyses in this appendix).

Outcome Types

Defining Outcome Types for Multilateral Investment Guarantee Agency Guarantee Projects

Similar to IFC projects, *RAP 2021* developed a 13-category typology of intended outcomes for MIGA guarantee projects by leveraging MIGA's Impact Measurement and Project Assessment Comparison Tool (IMPACT), the ex-ante assessment and monitoring tool that was adapted from IFC's AIMM system. No IMPACT sector-specific frameworks had been developed, so the same outcome typologies developed for IFC projects were applied in *RAP 2021*, with adaptations to some outcome types. Because MIGA had not retroactively applied IMPACT to its portfolio, the text of the president's reports on MIGA guarantee projects for coding of expected development outcomes were used for *RAP 2021*.

RAP 2023 again leveraged the outcome typology developed by *RAP 2021* and applied it to all MIGA guarantee projects that were self-evaluated by MIGA and validated by IEG between FY20 and FY22, but only those projects with PERs validated by December 2022 were included in the analysis. Because no IMPACT assessments were available for the projects in this cohort, the text of the "Development Outcome at Approval" section of IEG ValNotes was reviewed for *RAP 2023*, including the expected development outcome and coded descriptions of the project-level and foreign investment effects–level development outcomes the projects were intended to achieve. Furthermore, the outcome typology of *RAP 2021* was enhanced in *RAP 2023* by adding new categories and revising the definitions of some categories. *RAP 2023* identified 30 outcome types (25 project level and 5 foreign investment level; table A.16). Because IEG performed the coding manually, there is risk of subjective

assignment of outcome types for specific outcome claims. IEG shared its outcome analysis methodology with MIGA to gain feedback and ensure harmonization of views on the expected development outcomes of each MIGA guarantee project.

Methodology for Analysis on MIGA Guarantee Project Outcomes Achieved

For MIGA guarantee projects, *RAP 2023* followed the same approach to assess outcomes achieved as for IFC projects. Because no projects in the *RAP 2023* cohort had an IMPACT assessment, *RAP 2023* assessed the extent to which expected outcome claims were achieved at evaluation by verifying the results presented in the project ValNote. An outcome claim was considered fully achieved, partially achieved, not achieved, or cannot be verified based solely on the text of the project ValNote, which itself validated the project's self-evaluation PER. The *RAP 2023* team did not apply additional judgment, assessment, or methodology.

Table A.16. Outcome Typology for MIGA Guarantee Projects

Outcome Type	Description
1.1 - Access to goods and services	Increase in number of final beneficiaries of goods and services of the project or project enterprise; increase in volume of goods and services produced by project or project enterprise
1.1.1 - Access to goods and services (MSME)	Increase in number of MSMEs as final beneficiaries of goods and services of the project or project enterprise; increase in volume of goods and services produced or provided by project or project enterprise
1.1.2 - Access to goods and services (female)	Increase in number of final female beneficiaries of goods and services of the project or project enterprise; increase in volume of goods and services produced or provided by project or project enterprise
1.1.3 - Access to goods and services (customers)	Increase in number of individual customers as final beneficiaries of goods and services of the project or project enterprise; increase in volume of goods and services produced or provided by project or project enterprise

(continued)

Outcome Type	Description
1.1.4 - Access to goods and services (miscellaneous)	Increase in number of final beneficiaries of goods and services of the project or project enterprise other than MSMEs, female beneficiaries, and individual customers, or a mix of these final beneficiaries; increase in volume of goods and services produced by project or project enterprise
1.1.5 - Access to goods and services (project enterprise level)	Increase in capacity of the project or project enterprise to produce goods and services because of MIGA-guaranteed investment
1.2 – Quality and affordability of goods and services	Improved quality of goods and services produced by the project or project enterprise, compared with the baseline or with other producers or providers; lower production costs or process; reduced prices of goods and services, compared with the base-lines or other produces or providers
1.3 - Increased capacity of final beneficiaries	Enhanced capacity of final beneficiaries as a result of advisory services or training that is part of project scope
1.4 - Improved living standards (earnings) of individuals	Increase in revenue or decrease in expenditures by final beneficiaries (individuals) of goods and services produced by the project or project enterprise
1.5 - Improved sales or profitability of enterprises	Increase in revenue, decrease in expenditures, or increase in overall productivity by final beneficiaries (enterprises) of goods and services produced by project or project enterprise
1.6 - Economic return	Economic rate of return
1.7 - Financial and business performance of project enterprise	Financial and business performance of project enterprise, mostly project-executing agencies
2.1 - Suppliers and distributors reached	Increase in number of suppliers who provide inputs to the project or project enterprise, or expansion of network of distributors of goods or services produced by project or project enterprise
2.2 - Improved capacity of suppliers and distributors	Increase in capacity of suppliers or distributors as a result of advisory services or training that is part of project scope
2.3 - Improved sales and profitability of suppliers and distributors	Increase in volume of inputs provided by its suppliers or increase in the goods or services to be distributed by its distributors
3.1 - Increased employment	Increase in direct employment of project enterprise

(continued)

Outcome Type	Description
3.2 - Improved capacity and skills	Training provided to the employees of project or project enterprise
3.3 - Improved earning of employees	Increase in wages to employees of project or project enterprise
3.3 - Improved earnings of employees	Increase in wages to employees of project or company
4.1 - Increased transfers to government	Increase in payments by project or project enterprise to government, such as in the form of taxes, royalties, fees, or dividends
5.1 - Increased money spent or transferred to community	Increase in payments to communities around the project or project enterprise, such as on health, educational, vocational programs
6.1 - Enhanced environmental and social standards of the project enterprise	Improvement in environmental and social standards of project or project enterprise by MIGA
6.2 - GHG reduction	Decrease in or avoidance of greenhouse gas emissions
6.3 - Efficient use of resources	Decrease in use of water and other resources, improvement in solid waste management, implementation of waste-to-energy project
7.1 - Gross value added	Gross value added to economy (calculated based on a multiplier and expressed in monetary value)
7.2 - Induced or indirect employment	Induced and indirect employment based on multipliers
7.2 - Induced or indirect employment	Induced and indirect employment based on multipliers
7.3 - Export sales	Increase in exports of goods and services produced, generating foreign currency
8.1 - Governance	Improvement in corporate governance or increase in capacity of MIGA's guarantee project enterprise
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 through increased competitiveness, resilience, integration, enhancements to the regulatory environment, and so on

(continued)

Outcome Type	Description
11 - Development reach	Potential to stimulate future investments which increase inclusion and reduce inequality by reaching underserved populations (base of pyramid, women, youth, and so on)
12 - Sustainability	Potential to stimulate future investments to focus on climate change adaptation and mitigation, or adopting improved environmental and social standards and practices
13 - Signaling effects	Potential to stimulate further foreign investment in contexts where there are real or perceived barriers for domestic or foreign investors and lenders

Source: Independent Evaluation Group.

Note: MIGA = Multilateral Investment Guarantee Agency; MSME = micro, small, and medium enterprise.

[1] The Delivery Challenges in Operations for Development Effectiveness (DeCODE) taxonomy was developed by the World Bank's Global Delivery Initiative in 2016 and focuses on the typical delivery challenges that could impact project performance from design to closing. The taxonomy is comprehensive and well-structured, and its validation included a three-pronged iterative process comprising literature reviews, text analytics, and practitioners' consultations. It is structured at three levels of granularity: 3 clusters, 15 categories, and 52 subcategories.

[2] The original DeCODE training data set served solely as an initial reference to guide the coding stage of this *Results and Performance of the World Bank Group* (*RAP*) training set and to familiarize coders with the operationalization of various categories and subcategories as defined in the DeCODE taxonomy, but it was not used in the supervised machine learning exercise.

[3] Stop words" refer to words that are commonly present in a language and that do not typically contribute to the meaning of the document (for example, "a," "an," "the," "of," "in"). In addition to a standard corpus of stop words, the words "project" and "projects" and the phrases "World Bank" and "Bank Group" were removed from the vocabulary.

[4] Lemmatization refers to the process of reducing a word to its base form (for example, the lemmatized version of the word "better" is "good").

[5] Term frequency/inverse document frequency (TF-IDF) is a technique used to assign a numerical value to each word in a document to determine the importance of the word in the document

[6] Cosine similarity is a metric that quantifies the similarity between documents by measuring the inner angle between the vector representation corresponding to each documentt.

[7] Hyperparameters refer to configurations used to control the learning process.

[8] In addition to accuracy, the following metrics were calculated:

- Precision: The ratio of correct positive classifications (true positives) to the total number of positive results predicted by the classifier (true positives and false positives); weighted precision score on the testing set was 73 percent.
- Recall: the ratio of correct positive classifications (true positives) divided by the total positive samples (true positives and false negatives); weighted recall score on the testing set was 73 percent.

[9] Ensemble modeling is a process wherein multiple models are created to predict an outcome, for example, by using different modeling algorithms. The one-versus-rest strategy splits a multiclass classification into one binary classification problem per class. This approach was used only in cases in which the accuracy for a class was 50 percent or less. For example, assume that Class A has low accuracy, and Class B has the same or very similar accuracy as Class A. (The model cannot distinguish between these two classes very well.) Application of the one-versus-rest approach implies application of two additional logistic regression models: Class A versus all the other classes in the training set combined and Class B versus all the other classes in the training set combined. After this step, the performance of three different models must be compared (the main logistic regression model and the two new one-versus-rest models). To obtain the best performance, a common strategy in machine learning is to take the majority vote among these models. If two of these models make the same prediction, that prediction is maintained; otherwise, it is discarded. The goal of this approach is to increase confidence in the prediction of subcategories that the model could not clearly distinguish.

[10] The probabilities across all subcategories add up to 1.

[11] Projects closed before the start of the COVID-19 pandemic (March 2020) but with Implementation Completion and Results Reports completed after the start of the pandemic were excluded from the prepandemic set to avoid contaminating the prediction set because the team noticed that some of these projects included references to the pandemic in their narratives despite having closed before March 2020.

Appendix B. World Bank Project Rating and Bank Group Country Program Rating Trends and Patterns

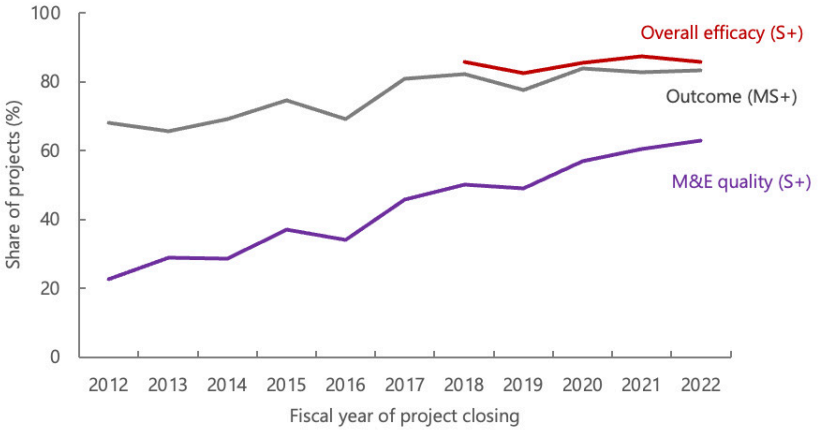
This appendix presents performance rating trends of World Bank projects and Bank Group country programs. Project performance rating trends cover lending operations validated and evaluated by IEG between fiscal year (FY)12 and FY22 through an Implementation Completion and Results Report Review and, in some cases, a Project Performance Assessment Report as of June 30, 2023, respectively. For projects that received both an Implementation Completion and Results Report Review and a Project Performance Assessment Report, the trend analysis includes only the latest evaluation ratings. The FY12–22 portfolio presented in the trend analysis includes 2,973 projects. Bank Group country program rating trends cover IEG's Completion and Learning Review Reviews (CLRRs) ratings for FY12-22.

Investment Project Financing and Program-for-Results

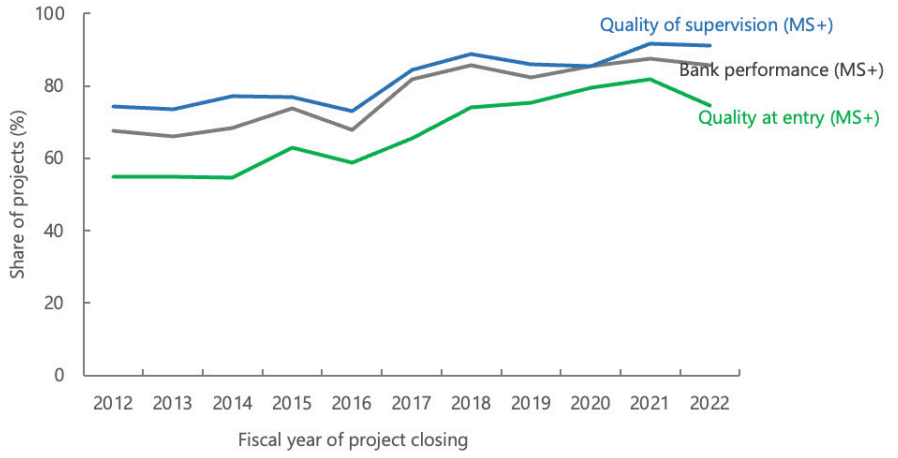
World Bank project outcome ratings remained high in FY22. The average outcome rating of 181 investment project financing and Program-for-Results projects in FY22 remained at 4.3 on a 6-point scale, which is the highest average since FY12 (figure B.2, panel b), with the share of projects rated moderately satisfactory or above staying constant (83 percent) between FY21 and FY22. There was a slight improvement in the share of investment project financing and Program-for-Results projects rated satisfactory or above, increasing from 47 percent in FY21 to 49 percent in FY22 (figures B.1, panel a, and B.2).

Figure B.1. World Bank Project Rating Trends: Investment Project Financing and Program-for-Results

a. Outcome, overall efficacy, and monitoring and evaluation quality



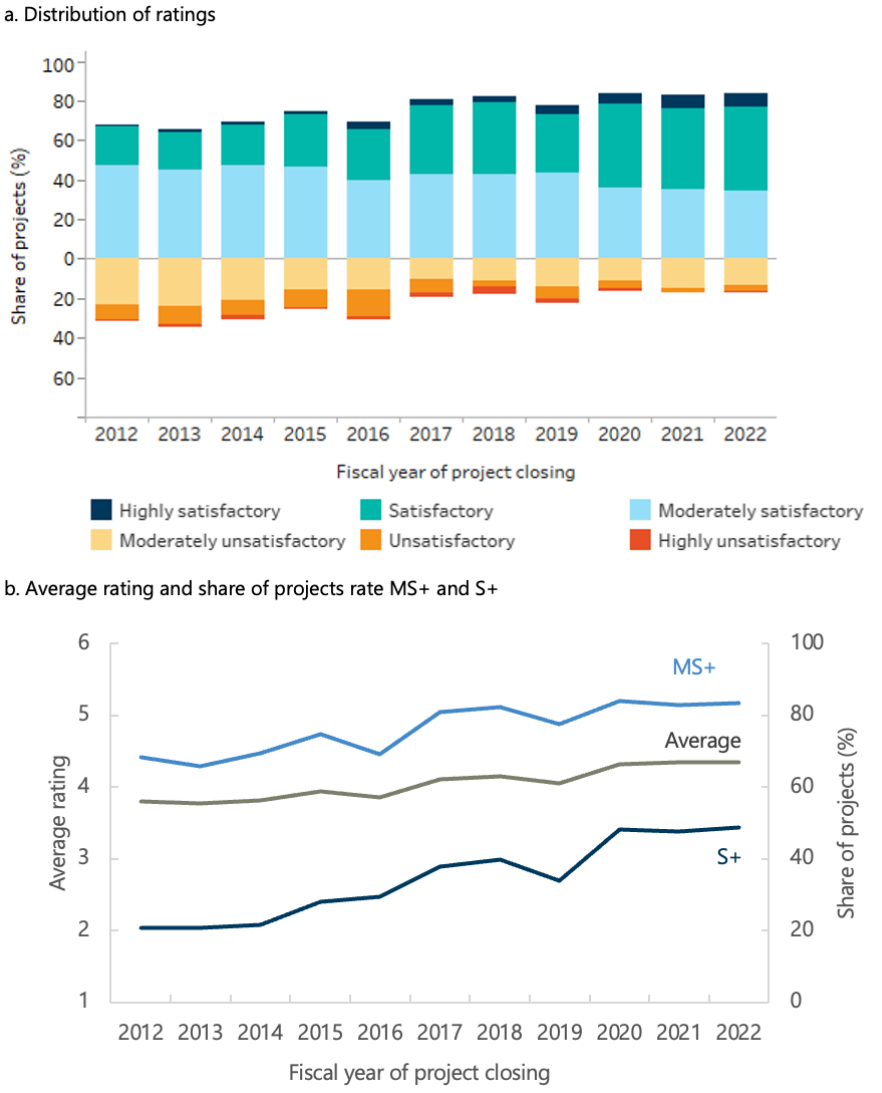
b. Bank performance, quality at entry, and quality of supervision



Source: Independent Evaluation Group.

Note: M&E = monitoring and evaluation; MS+ = moderately satisfactory or above; S+ = substantial or above.

Figure B.2. World Bank Project Outcome Rating: Investment Project Financing and Program-for-Results



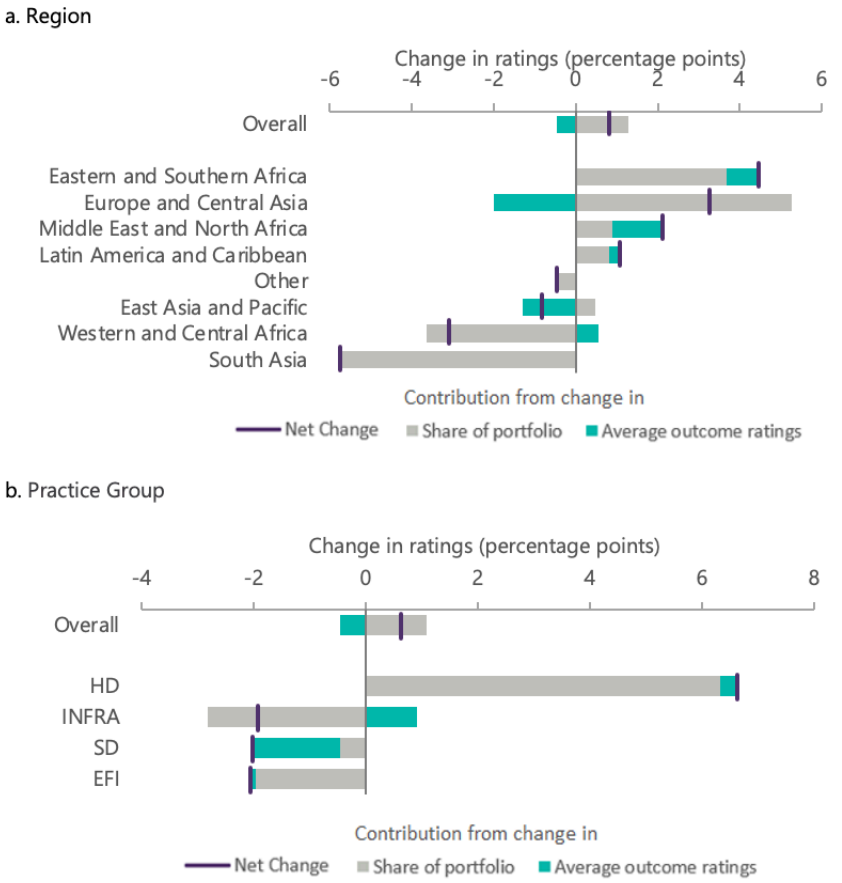
Source: Independent Evaluation Group.

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

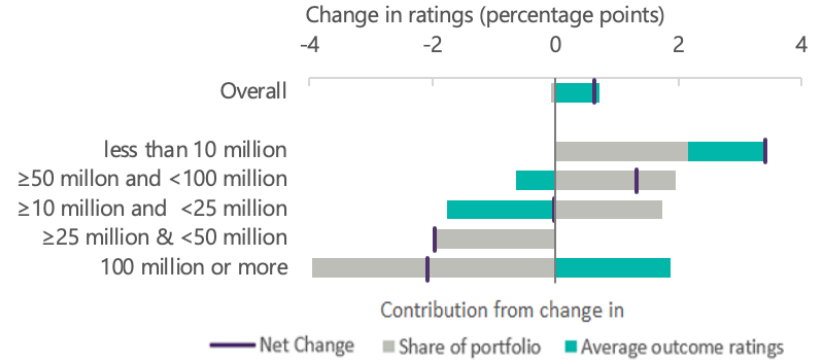
The decomposition analysis reveals that changes in ratings and portfolio allocations across different groups played a role in maintaining a consistent overall outcome rating from FY21 to FY22. Notably, Eastern and Southern Africa and the Human Development Practice Group had the greatest increase in contribution, driven by an improvement in ratings and an expanded

portfolio share. In addition, International Development Association countries not classified as fragile and conflict-affected situations increased the overall outcome rating, primarily because their portfolio share expanded. In contrast, South Asia's contribution declined, primarily because its portfolio share decreased. The Sustainable Development and Equitable Growth, Finance, and Institutions Practice Groups and International Bank for Reconstruction and Development countries not classified as fragile and conflict-affected situations also decreased the overall rating because their ratings and portfolio shares fell (figure B.3).

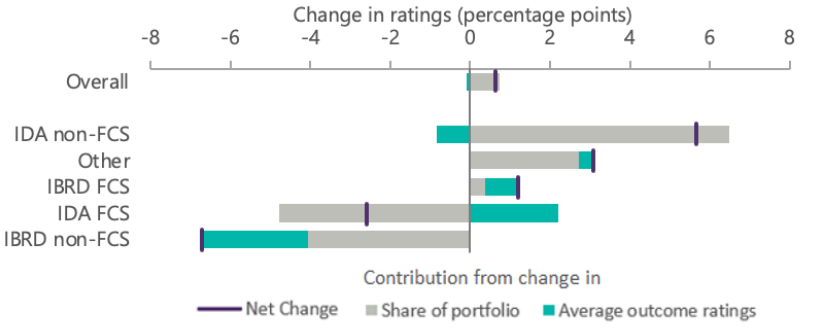
Figure B.3. Contributions to Shift in World Bank Project Outcome Ratings between Fiscal Years 2021 and 2022: Investment Project Financing and Program-for-Results



c. Project size



d. Lending group and fragile and conflict-affected situation status

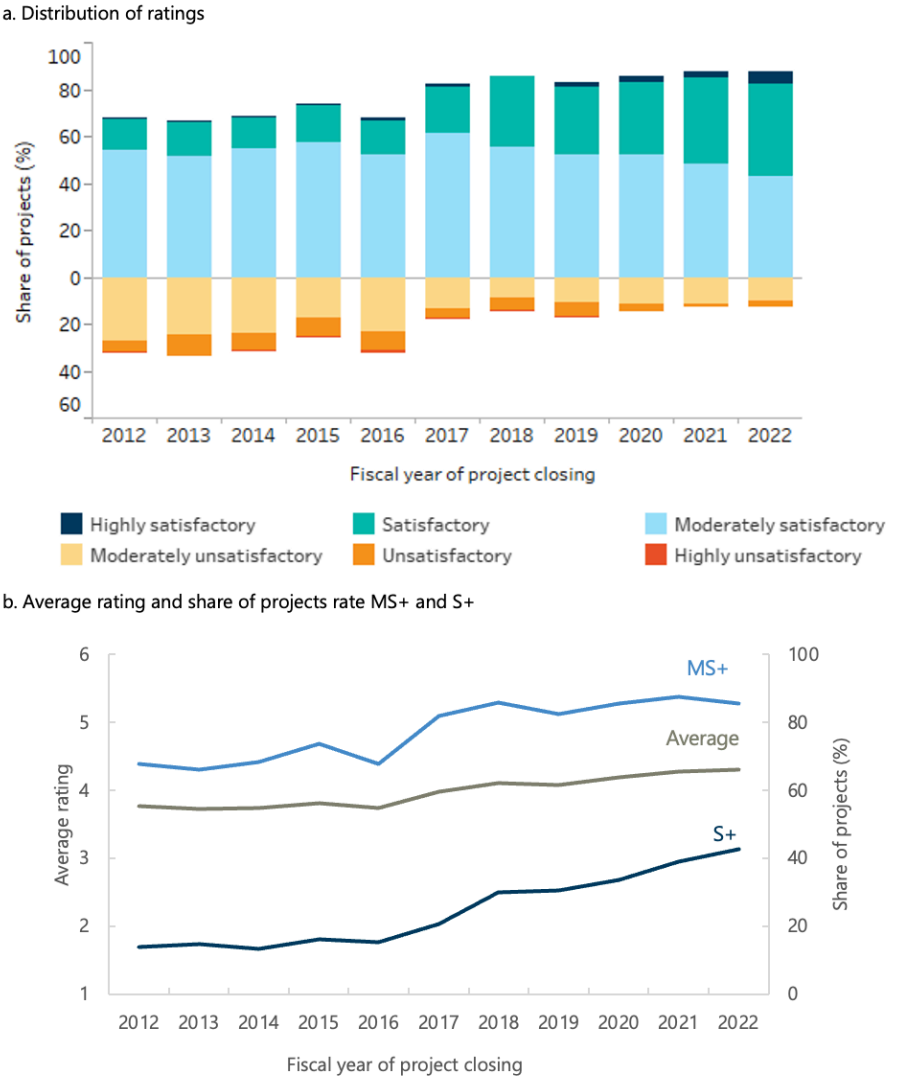


Source: Independent Evaluation Group.

Note: EFI = Equitable Growth, Finance, and Institutions; FCS = fragile and conflict-affected situation; HD = Human Development; IBRD = International Bank for Reconstruction and Development; IDA = International Development Association; INFRA = Infrastructure; SD = Sustainable Development.

Bank performance ratings for investment project financing and Program-for-Results projects also stayed flat, with an average rating of 4.3 on a 6-point scale in FY21 and FY22, which is the highest average since FY12. Although the share of projects rated moderately satisfactory or above declined marginally (from 87 percent in FY21 to 86 percent in FY22), the share of projects rated satisfactory or above increased from 39 percent in FY21 to 43 percent in FY22 (figures B.1, panel b, and B.4).

Figure B.4. World Bank Project Bank Performance Rating: Investment Project Financing and Program-for-Results



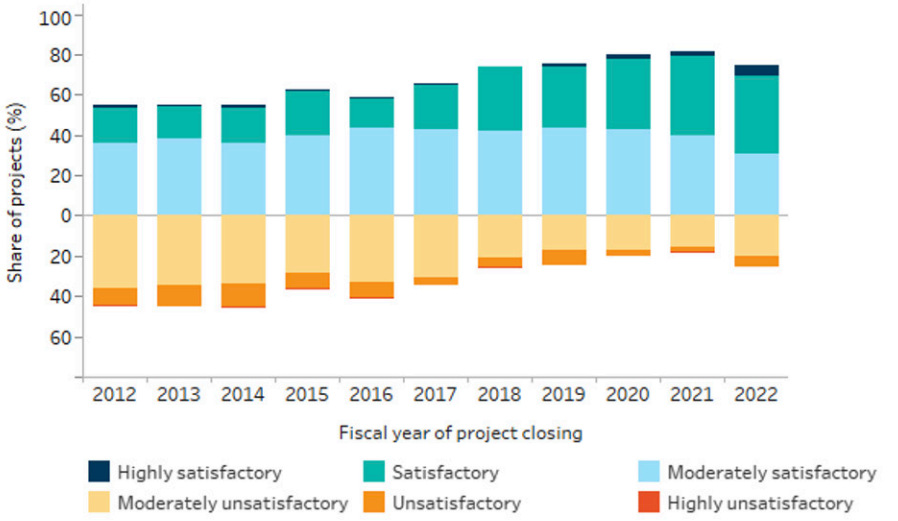
Source: Independent Evaluation Group.

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

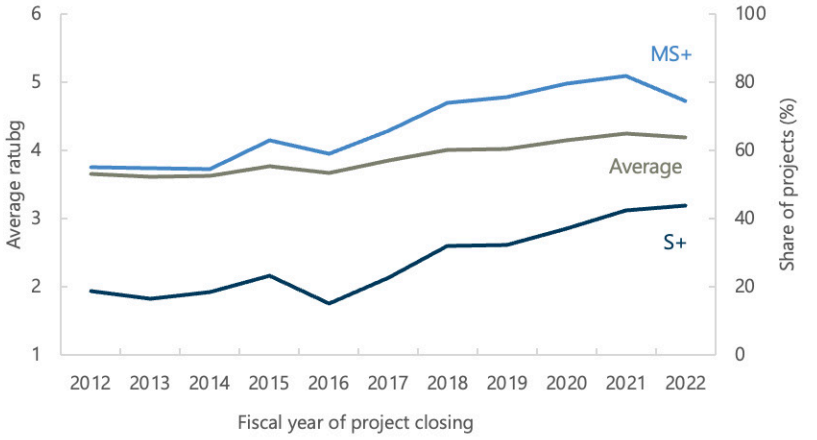
Average quality at entry, a subcomponent of Bank performance ratings, also remained constant at 4.2 on a 6-point scale, with an increase from 42 percent of projects rated satisfactory and above in FY21 to 44 percent in FY22 but also a decrease from 82 percent of projects rated moderately satisfactory and above in FY21 to 75 percent in FY22 (figure B.5).

Figure B.5. World Bank Project Quality at Entry Rating: Investment Project Financing and Program-for-Results

a. Distribution of ratings



b. Average rating and share of projects rate MS+ and S+



Source: Independent Evaluation Group.

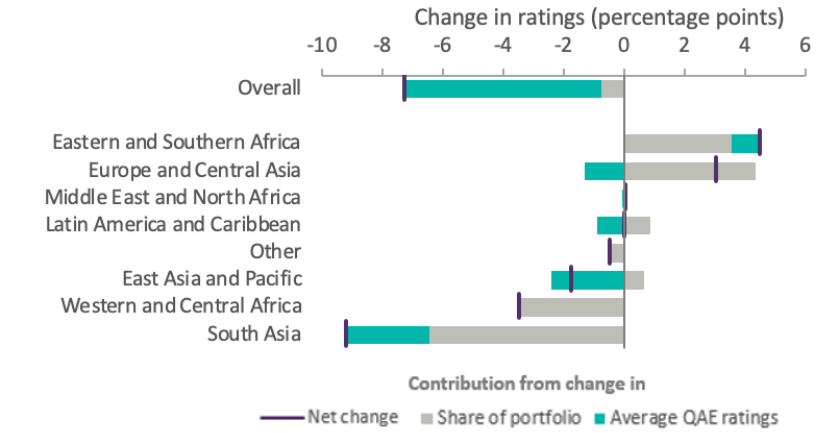
Note: MS+ = moderately satisfactory or above; S+ = satisfactory or above.

The decomposition analysis shows that, across World Bank Regions, the decline in quality at entry is mainly explained by the smaller portfolio share and sharp drop in project ratings in the South Asia Region, from 85 percent rated moderately satisfactory or above in FY21 to 60 percent in FY22. The decline in quality at entry ratings in FY22 were not linked to project

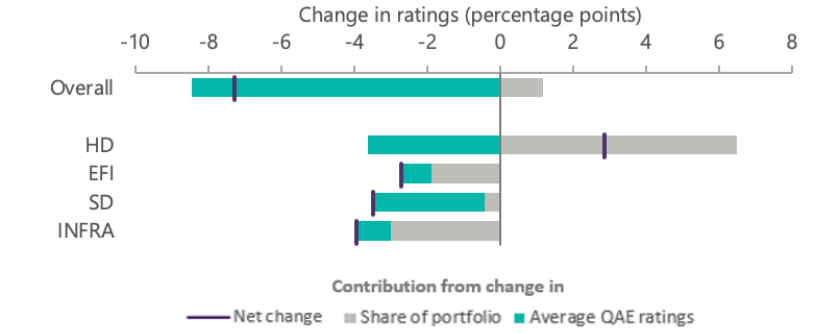
preparation challenges caused by COVID-19 because the vast majority of FY22 projects were approved before March 2020 (figure B.6).

Figure B.6. Contributions to Shift in World Bank Project Quality at Entry Ratings between Fiscal Years 2021 and 2022: Investment Project Financing and Program-for-Results

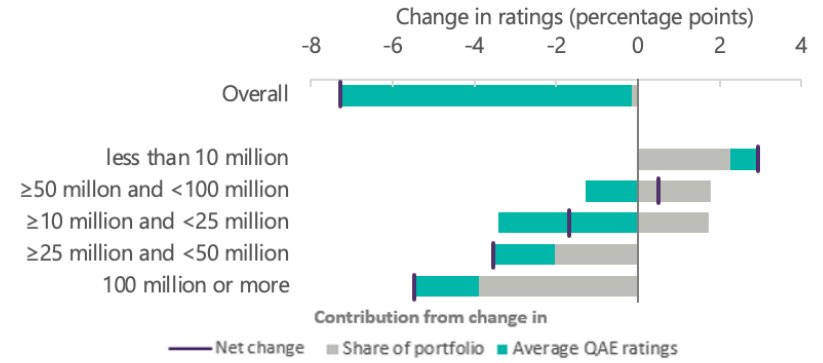
a. Region

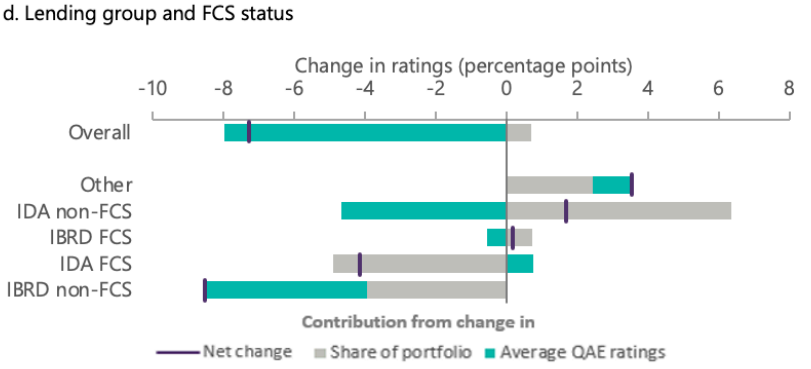


b. Practice Group



c. Project size



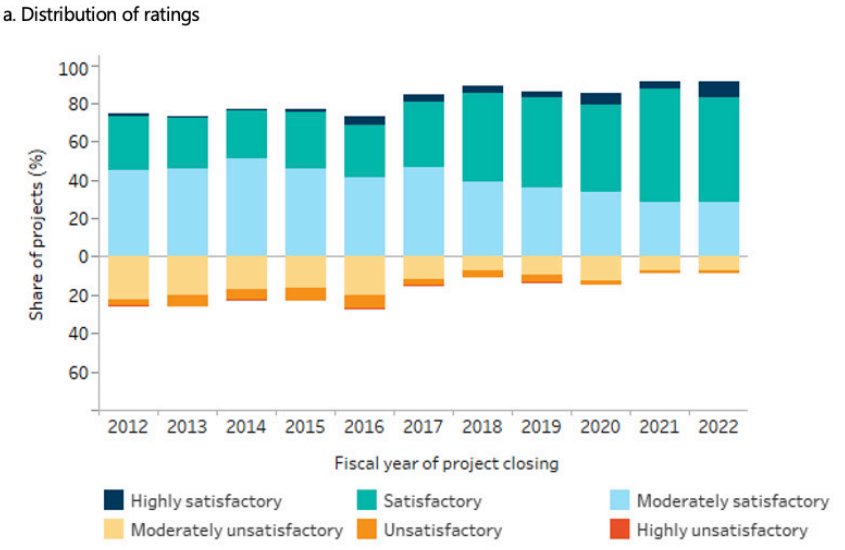


Source: Independent Evaluation Group.

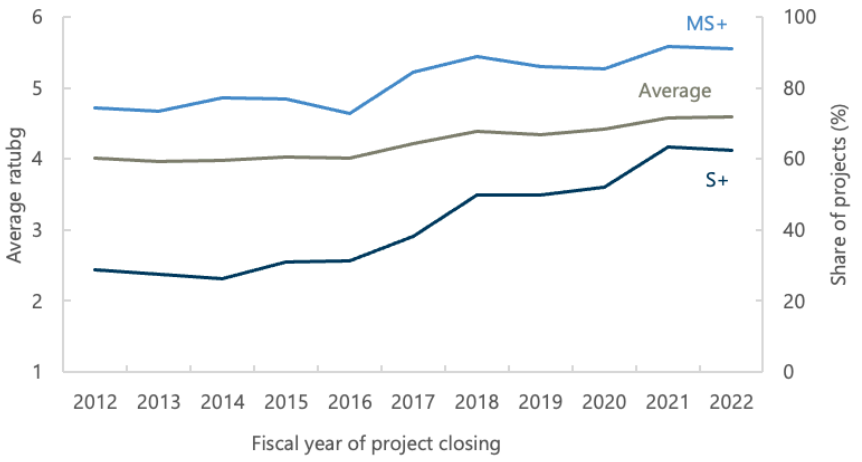
Note: EFI = Equitable Growth, Finance, and Institutions; FCS = fragile and conflict-affected situation; HD = Human Development; IBRD = International Bank for Reconstruction and Development; IDA = International Development Association; INFRA = Infrastructure; SD = Sustainable Development.

Quality of supervision, the other subcomponent of Bank performance ratings, also stayed constant, at 4.6 on a 6-point scale, with the share of projects rated highly satisfactory increasing from 4 percent in FY21 to 8 percent in FY22 and the share of projects rated moderately satisfactory or above slightly declining from 92 percent in FY21 to 91 percent in FY22 (figure B.7).

Figure B.7. World Bank Project Quality of Supervision Rating: Investment Project Financing and Program-for-Results



b. Average rating and share of projects rate MS+ and S+



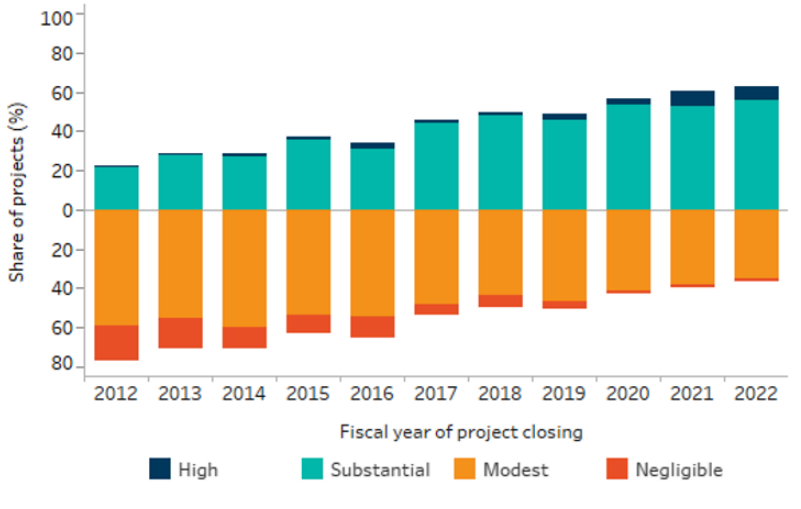
Source: Independent Evaluation Group.

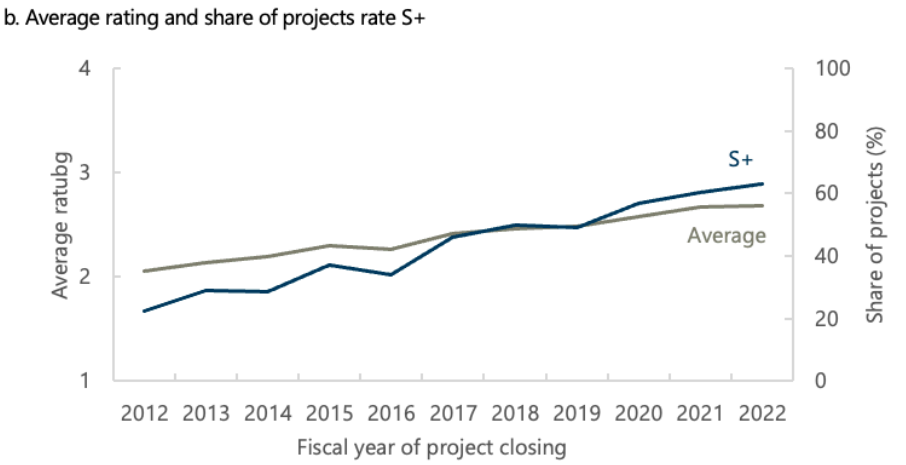
Note: MS+ = moderately satisfactory or above; S+ = satisfactory or above.

World Bank monitoring and evaluation quality ratings have consistently improved. The percentage of projects rated substantial or high in monitoring and evaluation quality increased from 60 percent in FY21 to 63 percent in FY22 (figure B.8).

Figure B.8. World Bank Project Monitoring and Evaluation Quality Rating: Investment Project Financing and Program-for-Results

a. Distribution of ratings





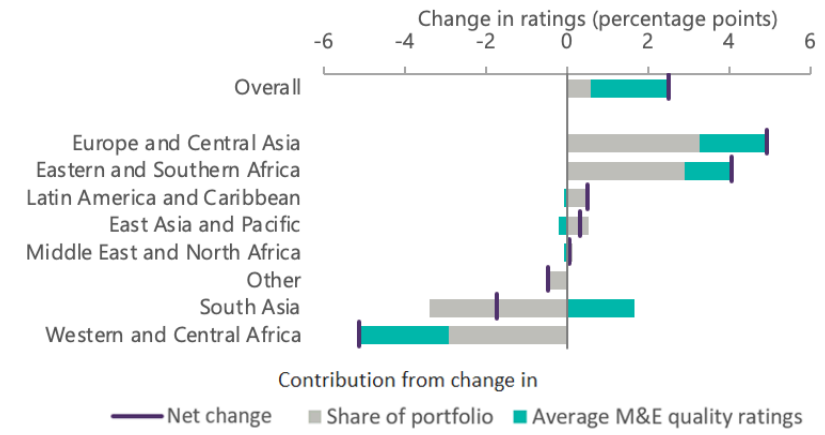
Source: Independent Evaluation Group.

Note: S+ = substantial or above.

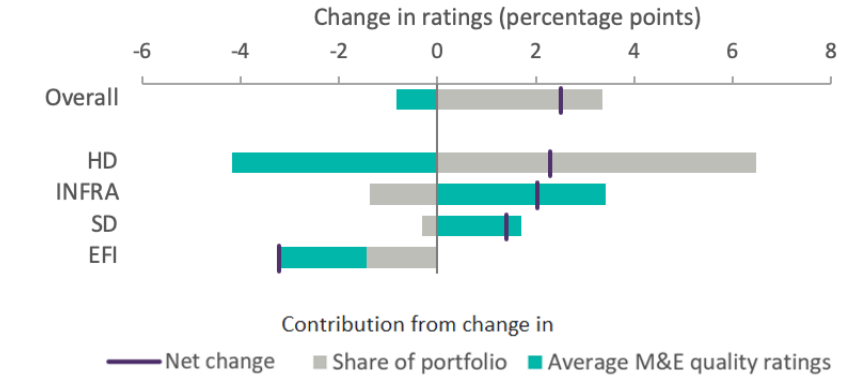
Decomposition analysis shows that the improved ratings of the Infrastructure Practice Group, from 37 percent to 56 percent, and the portfolio expansion of the high-performing Human Development Practice Group, from 20 percent to 27 percent of the overall portfolio, drove this increase. There was also a significant increase in monitoring and evaluation quality ratings in International Development Association countries classified as fragile and conflict-affected, from 48 percent to 60 percent. In regard to Regions, Western and Central Africa contributed negatively, with a decline in rating and portfolio share, whereas Europe and Central Asia and Eastern and Southern Africa drove the overall rating up, with an improvement in rating and a larger portfolio share (figure B.9).

Figure B.9. Contributions to the Shift in World Bank Project Monitoring and Evaluation Quality Ratings between Fiscal Years 2021 and 2022: Investment Project Financing and Program-for-Results

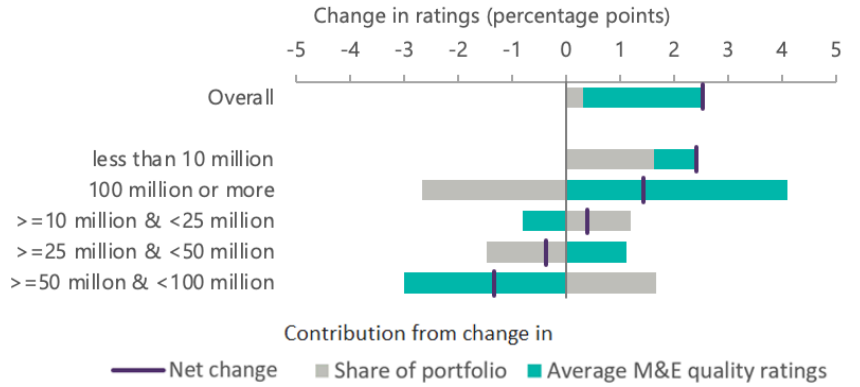
a. Region

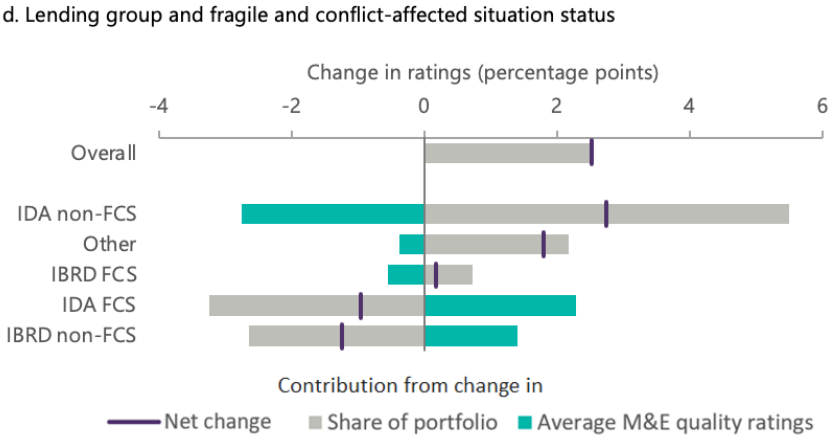


b. Practice Group



c. Project size





Source: Independent Evaluation Group.

Note: EFI = Equitable Growth, Finance, and Institutions; FCS = fragile and conflict-affected situation; HD = Human Development; IBRD = International Bank for Reconstruction and Development; IDA = International Development Association; INFRA = Infrastructure; M&E = monitoring and evaluation; SD = Sustainable Development.

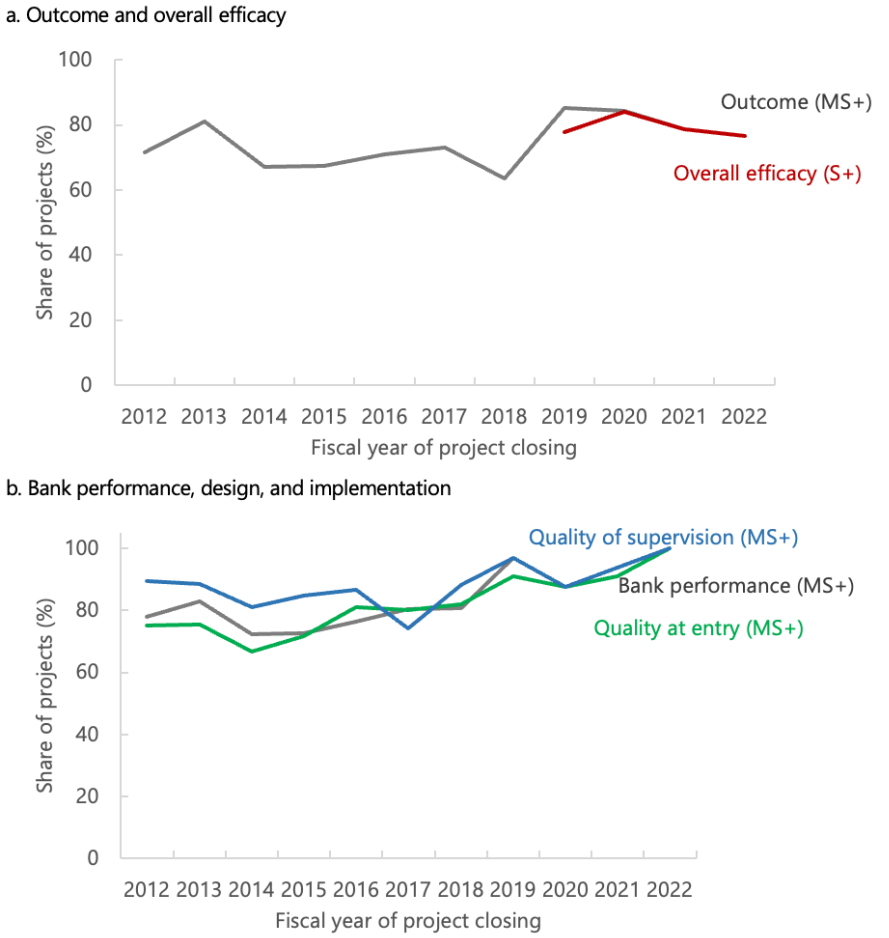
Development Policy Financing Projects

The average outcome rating for 17 development policy financing (DPF) projects in FY22 stayed at an average of 4.0 on a 6-point scale, with the percentage of projects rated satisfactory increasing slightly from 33 percent in FY21 to 35 percent in FY22. There was a fairly similar decline in the share of projects rated moderately satisfactory, from 45 percent in FY21 to 41 percent in FY22. Therefore, the share of DPF projects rated moderately satisfactory or above suffered a small decline from 79 percent to 76 percent (figures B.10-B.11).

Bank performance ratings for DPF projects improved from an average of 4.3 in FY21 to 4.6 in FY22 on a 6-point scale, and the percentage of projects rated moderately satisfactory or above increased from 94 percent in FY21 to 100 percent in FY22. Design and implementation ratings, which replaced quality at entry and quality of supervision ratings in DPFs, exhibited similar patterns, with the percentage of projects rated moderately satisfactory or above for design increasing from 91 percent in FY21 to 100 percent in FY22 and the percentage of projects rated moderately satisfactory or above for implementation ratings increasing from 94 percent in FY21 to 100 percent in FY22.

Performance ratings for DPF projects should be interpreted with caution because of limited sample size: 33 in FY21 and 17 in FY22.

Figure B.10. World Bank Project Rating Trends: Development Policy Financing

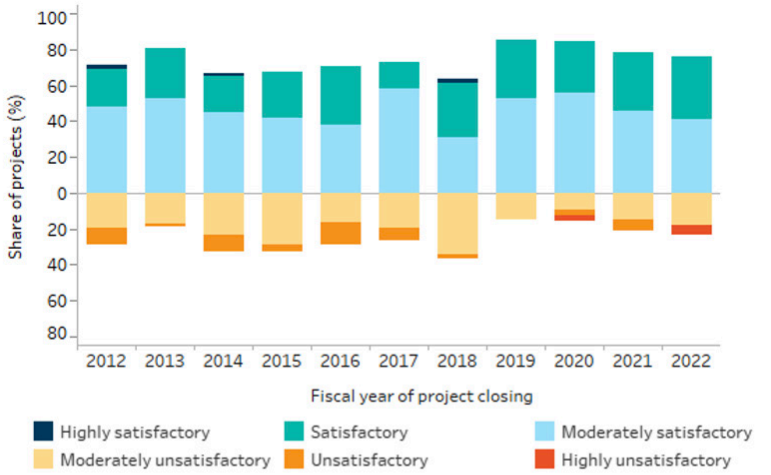


Source: Independent Evaluation Group.

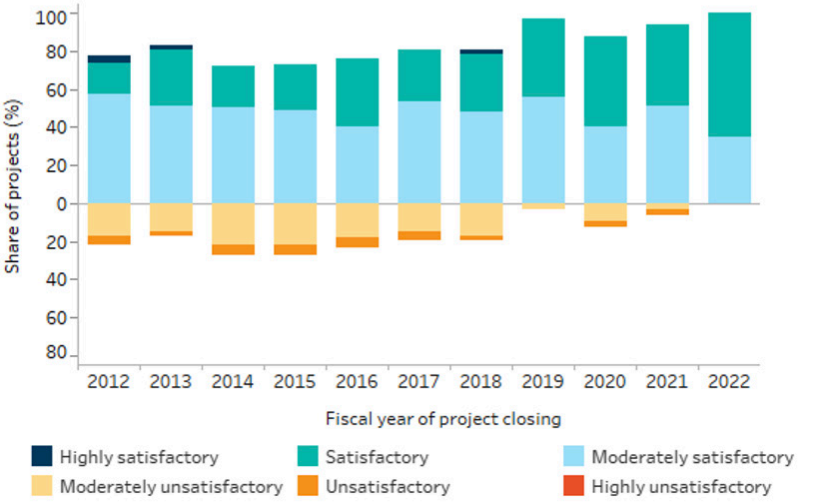
Note: MS+ = moderately satisfactory or above.

Figure B.11. Distribution of World Bank Project Ratings: Development Policy Financing

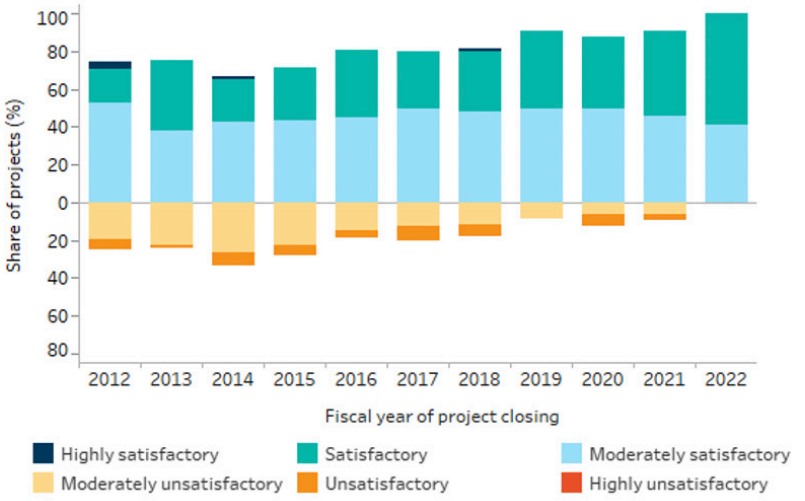
a. Outcome



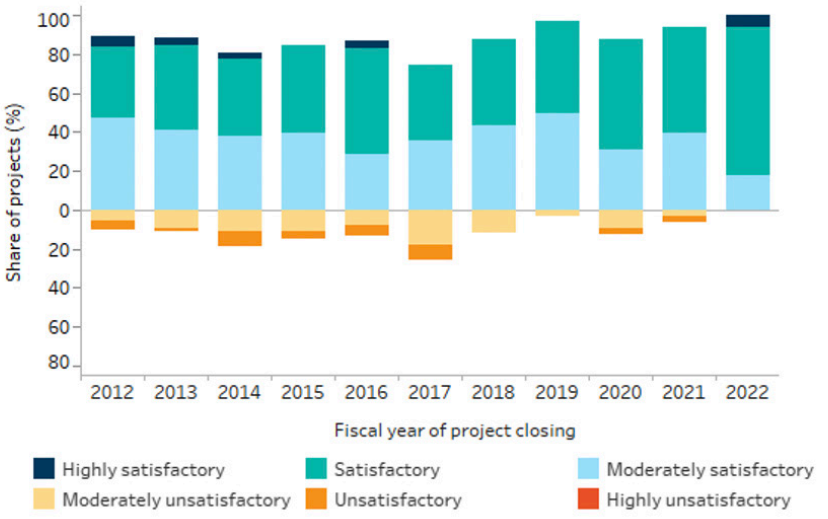
b. Bank performance



c. Design



d. Implementation



Source: Independent Evaluation Group.

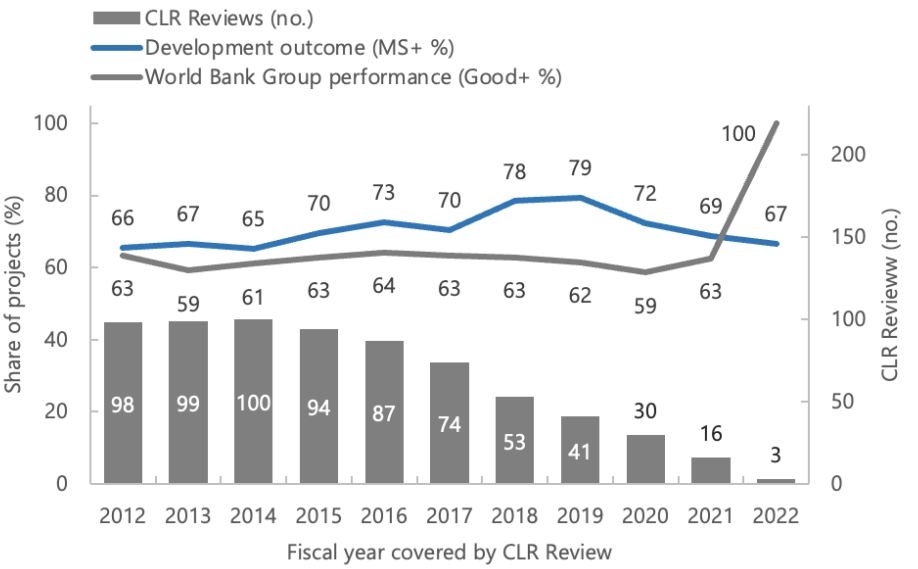
Performance of World Bank Group Country Programs

The upward trend in World Bank Group development outcome ratings, as observed in in the *RAP 2022*, has reversed in recent years. The share of development outcomes rated moderately satisfactory or above has declined since FY20, falling below the corporate target of 70 percent in FY22. However, it

is worth noting that the number of CLRRs covering recent fiscal years was small (30 in FY20, 16 in FY21, 3 in FY22; figure B.12).

The Bank Group performance rating declined moderately from FY16 to FY20 but improved in FY21. Except for FY22, which covers only 3 CLRRs, the rating has consistently remained below the corporate target of 75 percent since FY12 (figure B.12).

Figure B.12. Country Program Ratings, Fiscal Years 2012–22

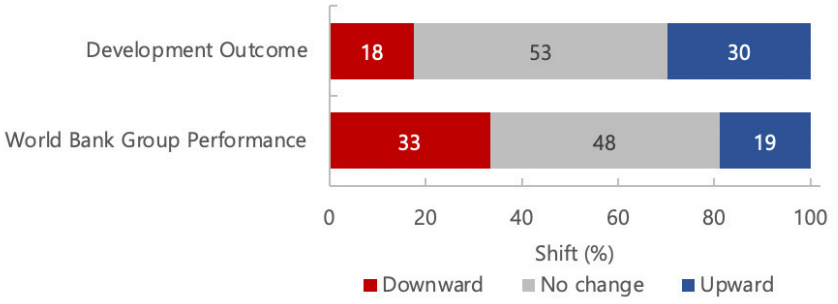


Source: Independent Evaluation Group.

Note: CLR = Completion and Learning Review.

To analyze the change in ratings over time, the analysis examined the two most recent CLRR ratings of countries with multiple CLRRs. The findings align with those of *RAP 2022*, indicating that the share of countries experiencing a downward trend in development outcome ratings is smaller than that of those showing an upward trend. Conversely, the share of countries experiencing a decrease in performance rating is larger than of those experiencing an increase (figure B.13).

Figure B.13. Shift in Development Outcome and World Bank Group Performance Ratings in Countries' Two Most Recent Completion and Learning Review Validations



Source: Independent Evaluation Group.

Appendix C. Analysis of World Bank Project Outcome Types

The outcome type analysis identifies the development outcomes that World Bank investment projects pursue in their development objectives. This appendix also explores the relationship between the intended development outcomes and objective-level efficacy ratings, considering the adequacy of the evidence provided in project results frameworks. Building on the principle of continuity that characterizes this *Results and Performance of the World Bank Group* (*RAP*), the analysis of outcome types expands on that in *RAP 2021*. (See appendix A for methodology.)

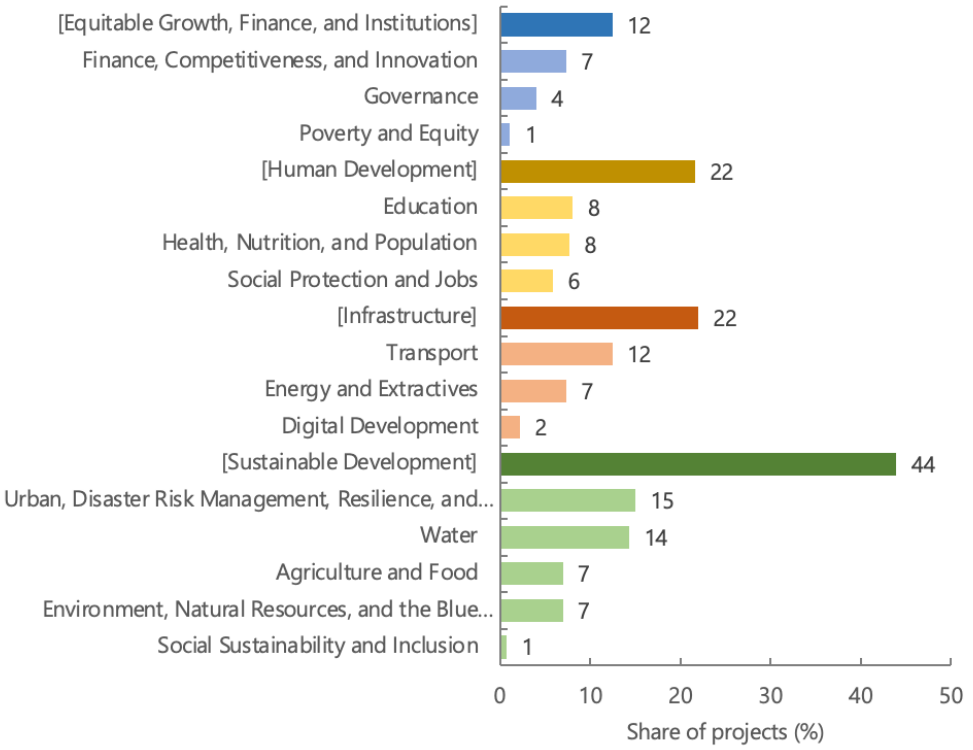
Results and Performance of the World Bank Group 2023 Cohort: A Portfolio Overview

The *RAP 2023* cohort comprises 273 projects evaluated during fiscal years (FY)20–22 that were approved during FY03–21.

Practice Groups and Global Practices. Of 273 total projects,

- 44 percent (120 projects) belong to the Sustainable Development Practice Group (Global Practices: Urban, Disaster Risk Management, Resilience, and Land; Water; Agriculture and Food; Environment, Natural Resources, and the Blue Economy; Social Sustainability and Inclusion; figure C.1);
- 22 percent (60 projects) belong to the Infrastructure Practice Group (Global Practices: Transport, Energy and Extractives; Digital Development); 22 percent (59 projects) to the Human Development Practice Group (Global Practices: Education; Health, Nutrition, and Population; Social Protection and Jobs); and
- 12 percent (34 projects) belong to the Equitable Growth, Finance, and Institutions Practice Group (Global Practices: Finance, Competitiveness, and Innovation; Governance; Poverty and Equity).

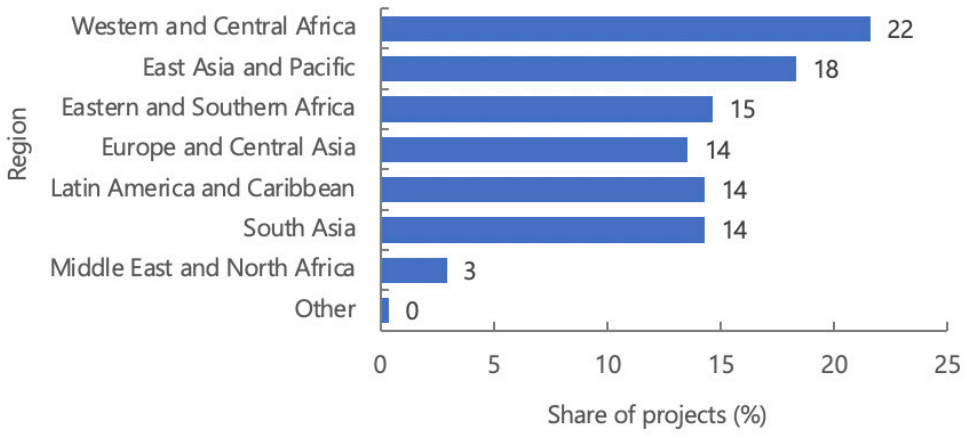
Figure C.1. Projects According to Practice Group and Global Practice



Source: Independent Evaluation Group.

Regions. Western and Central Africa has the largest share of projects (22 percent, 59 projects), and the Middle East and North Africa has the smallest (3 percent, 8 projects; figure C.2).

Figure C.2. Projects According to Region

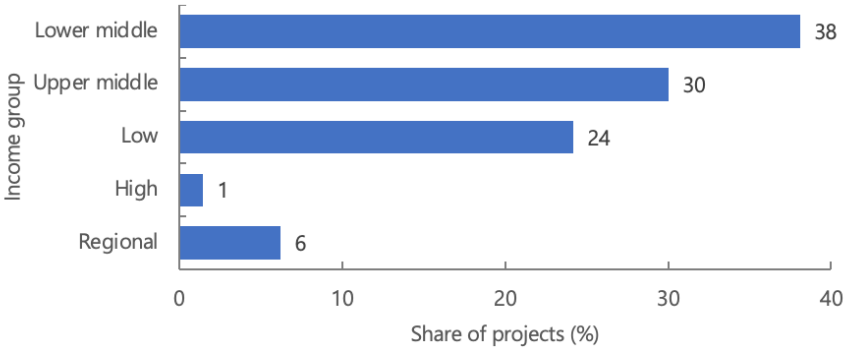


Source: Independent Evaluation Group.

Note: "Other" refers to projects tagged as "World."

Income groups. Most of the projects, 68 percent, are in middle-income countries. Within this category, 38 percent are in lower-middle-income countries, and 30 percent are in upper-middle-income countries (figure C.3).

Figure C.3. Projects According to Income Group

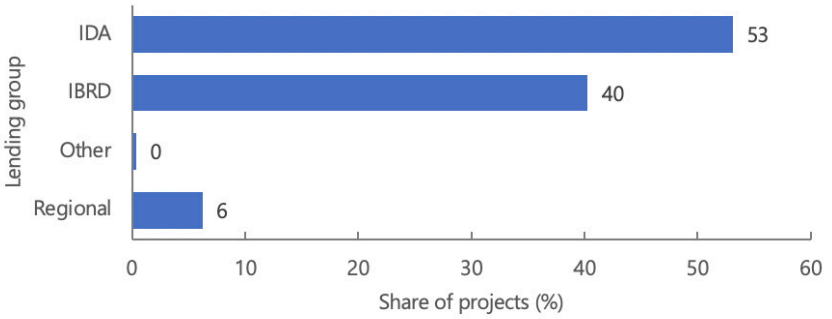


Source: Independent Evaluation Group.

Note: "Regional" refers to projects implemented in multiple countries.

Lending types. A significant portion of projects within the portfolio, 53 percent, are financed through credits and grants from the International Development Association Furthermore, 40 percent of the projects receive funding in the form of loans from the International Bank for Reconstruction and Development (figure C.4).

Figure C.4. Projects According to Lending Type

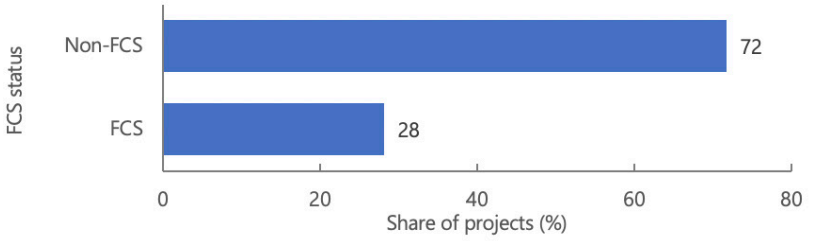


Source: Independent Evaluation Group.

Note: IDA = International Development Association; IBRD = International Bank for Reconstruction and Development. "Regional" refers to projects implemented in multiple countries.

Fragile and conflict-affected situations. Seventy-two percent of the portfolio (196 projects) is in non–fragile and conflict-affected situations (figure C. 5).

Figure C.5. Projects According to Fragile and Conflict-Affected Situation Status



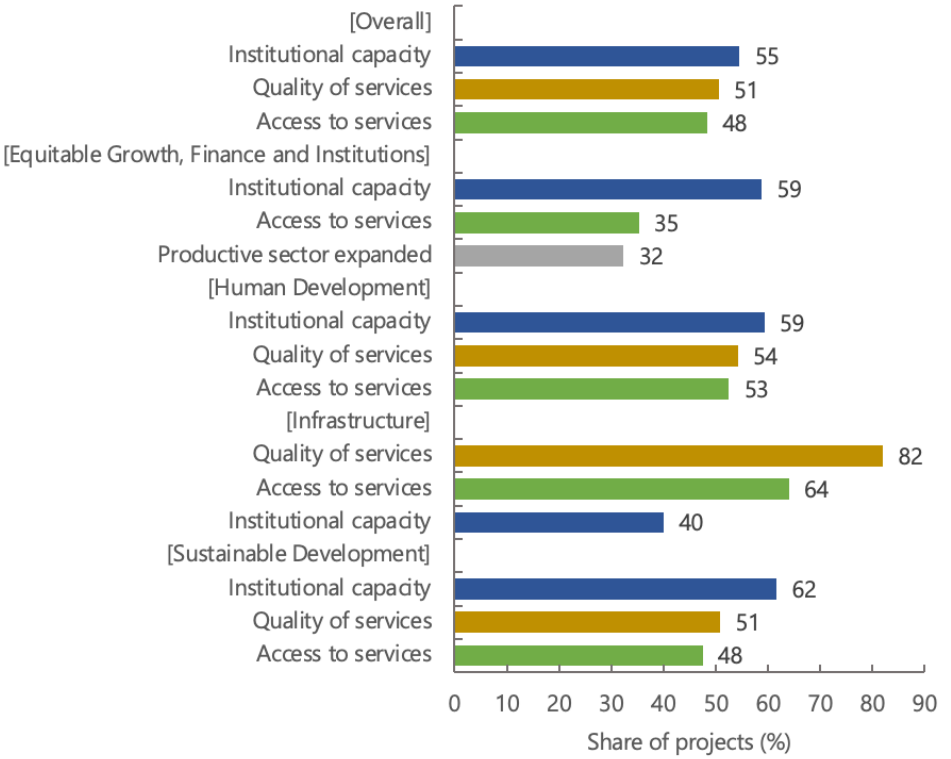
Source: Independent Evaluation Group.

Note: FCS = fragile and conflict-affected situation.

Development Outcome Types Pursued in the Results and Performance of the World Bank Group 2023 Cohort

The *RAP 2023* cohort focuses on three development outcome types: 55 percent of projects are designed to enhance the capacity of institutions to perform their institutional functions, 51 percent to improve the quality of services provided, and 48 percent to enhance access to services (figure C.6).[1] This aligns with the results reported in *RAP 2021*, which encompassed the FY12–14 and FY17–20 (second quarter) cohorts, highlighting the consistency in prioritization of these three outcome types.

Figure C.6. Top Three Development Outcomes per Practice Group



Source: Independent Evaluation Group.

These development outcome types are consistently pursued across various Practice Groups within investment projects, except for in the Equitable Growth, Finance, and Institutions Practice Group, in which expansion of

productive sectors is the third-most frequently pursued outcome type, accounting for 32 percent (table C.1).

There is noticeable variation in intended development outcomes across different Global Practices. Equity and inclusion outcomes were pursued in all projects under the Social Protection and Jobs and Social Sustainability and Inclusion Global Practices. The Agriculture and Food, Social Sustainability and Inclusion, and Social Protection and Jobs Global Practices commonly pursued individual employability and livelihood outcomes, at 58 percent, 50 percent, and 44 percent, respectively (table C.1).

Furthermore, certain development outcomes appeared to be more specific to particular Global Practices. For example, sustaining natural capital outcomes were the focus of 68 percent of the Environment, Natural Resources, and the Blue Economy projects. Increasing the use of services was a primary objective for 48 percent of the Health, Nutrition, and Population projects. Providing temporary relief to individuals was the primary goal for 63 percent of the Social Protection and Jobs projects. Citizen engagement and community participation development outcomes were prioritized in 50 percent of the Social Sustainability and Inclusion projects. Accountability and transparency objectives were more predominant in projects under the Poverty and Equity and Governance Global Practices.

There is not much variation in the top three outcomes across Regions (table C.2). However, there is more variation in outcome types when considering different country characteristics (table C.3). In the case of low-income countries, a notable emphasis is placed on increasing the capacity of institutions to perform institutional functions. For lower-middle-income countries, improving access to services emerges as an important outcome to pursue. Conversely, for both high-income and upper-middle-income countries, increasing natural capital becomes a significant outcome type to prioritize. This variation in outcome types across country characteristics highlights the nuanced approaches needed to address the specific development needs and priorities of different Regions and income groups.

Table C.1. Outcome Type According to Global Practice and Practice Group (percentage)

Outcome Type	All	EFI	FCI	GOV	POV	HD	EDU	HNP	SPL	INFRA	DDT	EAE	TDD	SD	AGR	ENV	SOC	URS	WAT
Institutional capacity	55	59	35	91	100	59	59	48	75	40	17	50	26	62	21	74	50	71	90
Quality of services	51	12	15	9	0	54	91	52	6	82	17	60	82	51	0	0	50	76	100
Access to services	48	35	50	18	0	53	41	52	69	64	83	40	56	48	26	0	50	76	69
Equity/inclusion	19	9	10	9	0	44	27	19	100	10	17	0	12	16	32	5	100	10	21
Individual employability/ livelihood	12	9	10	9	0	15	9	0	44	0	0	0	0	18	58	21	50	10	7
Natural capital	12	3	5	0	0	0	0	0	0	18	0	25	12	19	11	68	0	5	21
Use of services	8	0	0	0	0	19	5	48	0	16	33	5	15	3	0	0	0	2	7
Accountability/ transparency	7	29	10	45	100	10	27	0	0	4	0	10	0	2	0	0	0	5	0
Enterprise or sectoral performance	7	3	5	0	0	0	0	0	0	14	0	30	3	8	47	5	0	0	0
Productive sector expansion	7	32	55	0	0	0	0	0	0	4	17	5	0	4	16	0	0	5	0

(continued)

Outcome Type	All	EFI	FCI	GOV	POV	HD	EDU	HNP	SPL	INFRA	DDT	EAE	TDD	SD	AGR	ENV	SOC	URS	WAT
Awareness/ attitudes/ behaviors	5	3	5	0	0	2	0	5	0	2	17	0	0	8	11	21	0	0	14
Temporary relief to individuals	4	0	0	0	0	19	5	0	63	0	0	0	0	1	5	0	0	0	0
Legal or regulatory context	4	18	20	18	0	0	0	0	0	4	33	0	0	2	0	5	0	2	0
Human capital	2	3	5	0	0	5	9	0	6	2	0	0	3	0	0	0	0	0	0
Public assets	1	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	10	0
Citizen engagement/ community participation	1	3	0	9	0	0	0	0	0	0	0	0	0	3	0	5	50	2	0

Source: Independent Evaluation Group.

Note: EFI= Equitable Growth, Finance, and Institutions; FCI = Finance, Competitiveness, and Innovation; GOV = Governance; POV = Poverty and Equity; HD = HumanDevelopment; EDU = Education; HNP = Health, Nutrition, and Population; SPL = Social Protection and Jobs; INFRA = Infrastructure; DDT = Digital Development; EAE = Energy and Extractives; TDD = Transport; SD = Sustainable Development; AGR = Agriculture and Food; ENB = Environment, Natural Resources, and Blue Economy; SOC = Social Sustainability and Inclusion; URS = Urban, Disaster Risk Management, Resilience, and Land; WAT = Water.

Table C.2. Outcome Type According to Region (percentage)

Outcome Type	All	AFE	AFW	EAP	ECA	LCR	MNA	SAR	Other
Institutional capacity	55	65	54	38	59	59	38	62	0
Quality of services	51	55	39	60	59	38	50	56	0
Access to services	48	68	42	46	30	46	63	56	100
Equity/inclusion	19	15	25	12	11	31	13	23	0
Individual employability/ livelihood	12	10	15	10	5	15	13	18	0
Natural capital	12	10	12	16	5	15	25	10	0
Use of services	8	3	14	18	0	5	13	3	0
Accountability/ transparency	7	10	8	2	14	8	0	5	0
Enterprise or sectoral performance	7	3	7	8	8	10	0	5	0
Productive sector expansion	7	10	7	6	11	3	13	3	0
Awareness/ attitudes/ behaviors	5	0	3	4	0	8	13	13	0
Temporary relief to individuals	4	8	8	0	3	5	13	0	0
Legal or regulatory context	4	8	2	4	5	0	0	5	0
Human capital	2	0	0	4	3	3	0	3	0
Public assets	1	3	0	0	5	0	0	3	0
Citizen engagement/ community participation	1	0	5	2	0	0	0	0	0

Source: Independent Evaluation Group.

Note: AFE = Eastern and Southern Africa; AFW = Western and Central Africa; EAP = East Asia and Pacific; ECA = Europe and Central Asia; LCR = Latin America and the Caribbean; MNA = Middle East and North Africa; SAR = South Asia.

Table C.3. Outcome Type According to Country Income and Lending (percentage)

Outcome Type	All	Low Income	Lower-Middle Income	Upper-Middle Income	High Income	IDA Non-FCS	IDA FCS	IBRD Non-FCS	IBRD FCS	Regional
Institutional capacity	55	64	52	50	100	61	56	50	67	47
Quality of services	51	50	54	55	25	54	51	53	33	18
Access to services	48	44	55	44	0	51	49	44	33	59
Equity/inclusion	19	15	27	18	0	26	18	19	33	0
Individual employability/livelihood	12	18	13	11	0	14	14	13	0	0
Natural capital	12	6	11	21	25	11	4	20	33	0
Use of services	8	6	8	10	0	4	7	11	0	12
Accountability/transparency	7	14	5	6	25	4	15	6	0	0
Enterprise or sectoral performance	7	6	6	7	0	6	5	7	0	12
Productive sector expansion	7	8	5	7	0	7	7	6	0	12
Awareness/attitudes/behaviors	5	5	6	4	0	6	4	5	0	6
Temporary relief to individuals	4	8	5	2	0	10	4	1	33	0
Legal or regulatory context	4	6	1	4	0	1	5	3	0	12
Human capital	2	0	2	4	0	0	0	5	0	0

(continued)

Outcome Type	All	Low Income	Lower-Middle Income	Upper-Middle Income	High Income	IDA Non-FCS	IDA FCS	IBRD Non-FCS	IBRD FCS	Regional
Public assets	1	0	3	1	0	3	0	2	0	0
Citizen engagement/community participation	1	5	1	0	0	1	4	0	0	0

Source: Independent Evaluation Group.

Note: FCS = fragile and conflict-affected situation; IBRD = International Bank for Reconstruction and Development; IDA = International Development Association.

Results Framework Indicators in the Results and Performance of the World Bank Group 2023 Cohort

To assess the validity of the results frameworks to measure the intended development outcomes, this *RAP* examined 4,808 results framework indicators corresponding to the 273 projects included in the portfolio. This assessment aimed to determine whether the indicators effectively captured and measured the desired project development outcomes (see appendix A for methodology.)

About 30 percent of results framework indicators were project development objectives (PDO) indicators, and 70 percent were intermediate results indicators. PDO indicators are designed to measure the intended impact or outcome of the project and are usually limited in number. Intermediate results indicators are designed to measure the progress of specific components or activities of the project that contribute to achievement of the project's overall goal and tend to be more numerous than PDO indicators.

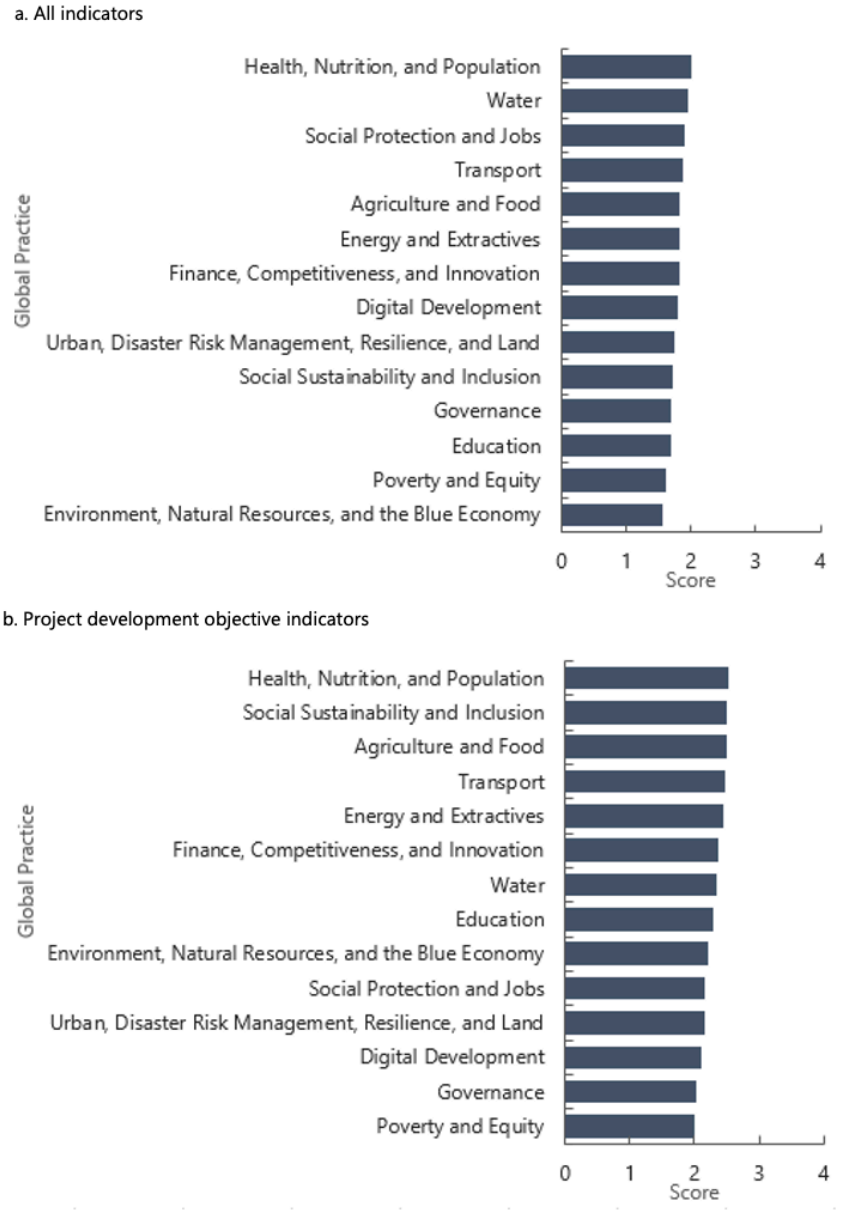
Outcome type matching. Our analysis found that the results framework indicators align closely with the intended development outcomes implied by the project objectives. More than 90 percent of the cases showed a match between the indicators and the desired development objectives outcomes.

Indicator level. PDO indicators predominantly focus on measuring outcomes (40 percent) and intermediate outcomes (46 percent), with 12 percent measuring outputs and 2 percent measuring high outcomes. Intermediate results indicators mostly focused on lower-level indicators, which are designed to assess the progress of specific project components or activities and offer operational insights into project progress. They mostly measure outputs (53.6 percent), followed by intermediate outcomes (38.2 percent), outcomes (8.2 percent), and high outcomes (0.1 percent).

When assessed on a 4-point scale, project objectives have an average indicator-level score of 2.3 as measured by PDO indicators, meaning that most indicators are at the outcome and intermediate outcome levels. Notably, project objectives within the Health, Nutrition, and Population Global

Practice have the highest indicator-level score (average 2.5), corresponding to the intermediate outcome and outcome levels and suggesting an emphasis on measuring outcomes beyond mere outputs (figure C.7).

Figure C.7. Average Indicator-Level Score According to Objective and Global Practice

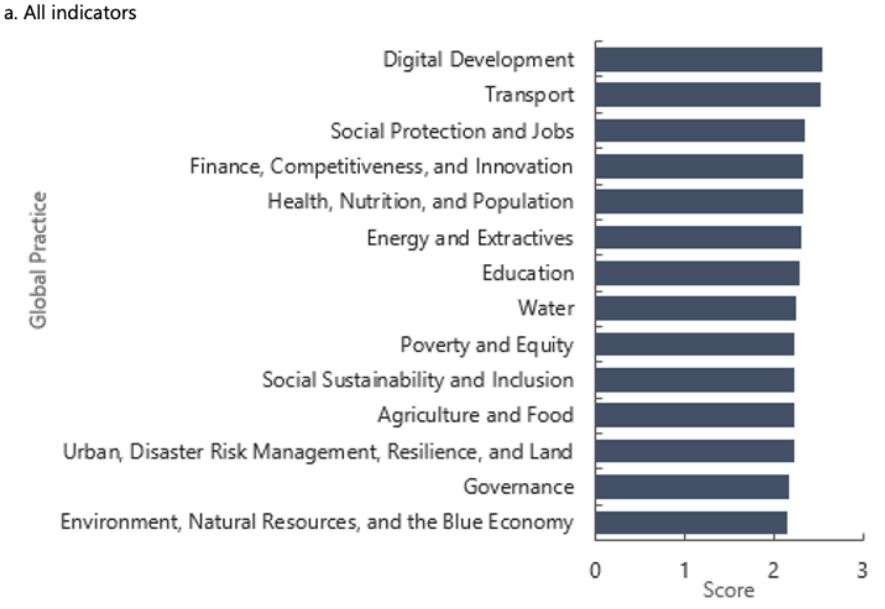


Source: Independent Evaluation Group.

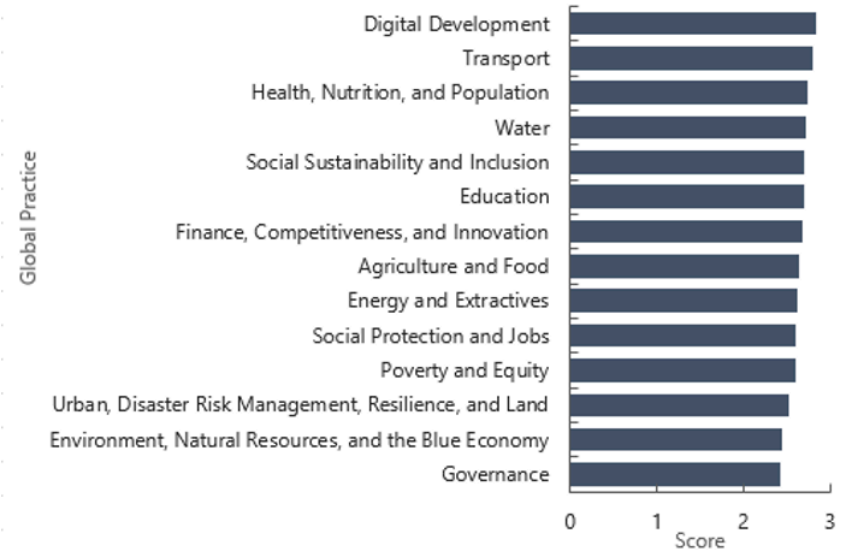
Note: Indicator level is measured on a four-point scale: 1 = output, 2 = intermediate outcome, 3 = outcome, 4 = high outcome.

Indicator adequacy. Given the distinction between PDO and intermediate results indicators in a given project, not all indicators included in the results framework are sufficient to demonstrate achievement of the development objective. However, they are important for providing evidence of completion of project activities based on the project's theory of change. In the *RAP 2023* cohort, 85 percent of development objectives had at least one fully adequate indicator to measure the intended development outcome. When measured on a 3-point scale, the average indicator adequacy score was 2.3. This suggests a combination of partially and fully adequate indicators in capturing the intended development outcomes. It is worth noting that the Digital Development and Transport Global Practices had slightly higher average adequacy scores than the other Global Practices (figure C.8), which indicates a relatively stronger alignment between the indicators and the intended outcomes in these areas.

Figure C.8. Average Indicator Adequacy Score According to Objective and Global Practice



b. Project development objective indicators



Source: Independent Evaluation Group.

Note: Indicator adequacy is measured on a 3-point scale: 1 = not adequate, 2 = partially adequate, 3 = fully adequate.

Project Performance Ratings of the Results and Performance of the World Bank Group 2023 Cohort

Correlation across project performance ratings. Table C.4 indicates strongly positive and significant associations between project performance ratings, emphasizing the high consistency and interconnectedness in assessing project performance across various dimensions.

Table C.4. Correlation Matrix of Project Ratings in the *RAP 2023* Cohort (Pearson correlation coefficients)

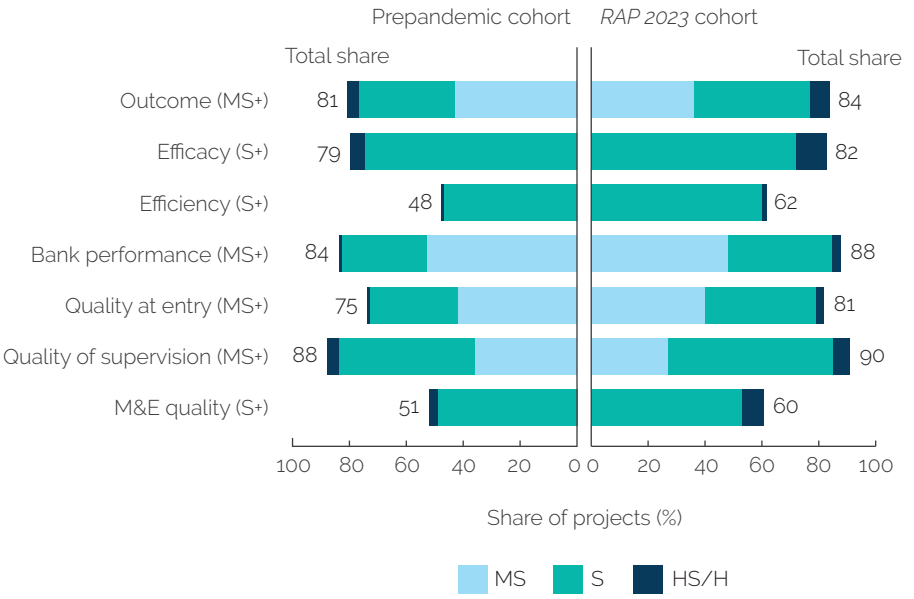
	Overall Efficacy	Outcome	Bank Performance	Quality at Entry	Quality of Supervision	M&E Quality
Overall efficacy	1.0***	0.85***	0.67***	0.57***	0.62***	0.54***
Outcome	0.85***	1.0***	0.77***	0.68***	0.69***	0.57***
Bank performance	0.67***	0.77***	1.0***	0.88***	0.81***	0.59***
Quality at entry	0.57***	0.68***	0.88***	1.0***	0.68***	0.59***
Quality of supervision	0.62***	0.69***	0.81***	0.68***	1.0***	0.56***
M&E quality	0.54***	0.57***	0.59***	0.59***	0.56***	1.0***

Source: Independent Evaluation Group.

Note: M&E = monitoring and evaluation; *RAP* = *Results and Performance of the World Bank Group*.
***$p < .01$.

Improved project performance ratings. Projects in the *RAP 2023* cohort performed better than those in the prepandemic cohort. A comparative analysis of project performance ratings between cohorts demonstrates statistically significantly higher ratings for projects that were exposed to the pandemic. This finding is supported by figure C.9 and table C.5, which provide a comparison of ratings between the two cohorts. This suggests that despite the implementation challenges facing projects (as outlined in appendix D), project teams were able to adapt and effectively navigate through the obstacles presented by the pandemic and other unfavorable circumstances.

Figure C.9. World Bank Project Ratings: Prepandemic (FY18–20) versus *RAP 2023* (FY20–22) Cohorts



Source: Independent Evaluation Group.

Note: FY = fiscal year; H = high; HS = highly satisfactory; M&E = monitoring and evaluation; MS = moderately satisfactory; *RAP* = *Results and Performance of the World Bank*; S = substantial or satisfactory; .

Table C.5. Average Project Ratings between FY18–20 and FY20–22

			Student t-Test	Mann-Whitney U Test
Rating	Prepandemic	*RAP 2023* Cohort	*p* value	
Outcome	4.11	4.36	0.00	0.00
Efficacy	2.81	2.91	0.02	0.03
Efficiency	2.42	2.62	0.00	0.00
Monitoring and evaluation quality	2.49	2.66	0.00	0.00

(continued)

Rating	Prepandemic	*RAP 2023* Cohort	Student t-Test	Mann-Whitney U Test
			p value	
Bank performance	4.11	4.29	0.01	0.01
Quality at entry	4.03	4.23	0.00	0.01
Quality of supervision	4.39	4.58	0.00	0.00

Source: Independent Evaluation Group.

Note: Outcome, Bank performance, quality at entry, and quality of supervision ratings are rated on a six-point scale. Efficacy, efficiency, and monitoring and evaluation quality ratings are rated on a four-point scale. FY = fiscal year; RAP = Results and Performance of the World Bank Group.

The improvement in project performance ratings in the *RAP 2023* cohort is not indicative of a systematic difference from previous years in the type of development outcomes pursued. The top three development outcomes pursued in World Bank projects, as observed in *RAP 2021* across FY12–14 and FY17–20 (second quarter), continue to be increasing institutional capacity, improving service quality, and expanding access to services.

Furthermore, the distribution of intended development outcomes pursued over time was relatively constant across the cohorts. The efficacy ratings have shown consistent improvement, and this upward shift is statistically significant in the long run. The comparison between FY12–14 and FY17–20 (second quarter) and between FY12–14 and FY20–22 demonstrates this significant increase in efficacy (table C.6) and indicates the World Bank's ongoing efforts to enhance project effectiveness, which is reflected in the efficacy ratings during FY20–22 (figure C.10).

Table C.6. Outcome Types and Average Efficacy Ratings: FY12–14, FY17–20(Q2), FY20 (March)–22

Outcome Type	Share of Objectives (%)			Average Efficacy Rating					
	FY12–14	FY17–20(Q2)	FY20 (March)–22	FY12–14		FY17–20 (Q2)		FY20 (March)–22	
Access to services expanded	23	25	30	2.63	*a,y*	2.85	*a*	2.87	*y*
Quality of services improved	40	47	36	2.59	*a,y*	2.77	*a*	2.83	*y*
Public assets improved	1	3	1	2.33		2.73		3.00	
Natural capital sustained	8	8	7	2.44	*y*	2.62		2.89	*y*
Use of services of assets increased	6	7	5	2.55	*y*	2.76		2.94	*y*
Temporary relief to individuals provided	3	2	3	2.64		3.00		2.89	
Awareness, attitudes, or behaviors changed	4	4	3	2.79		2.74		2.53	
Human capital increased	2	7	1	2.60		2.82		2.78	
Individual employability or livelihood improved	4	4	7	2.55		2.65		2.64	

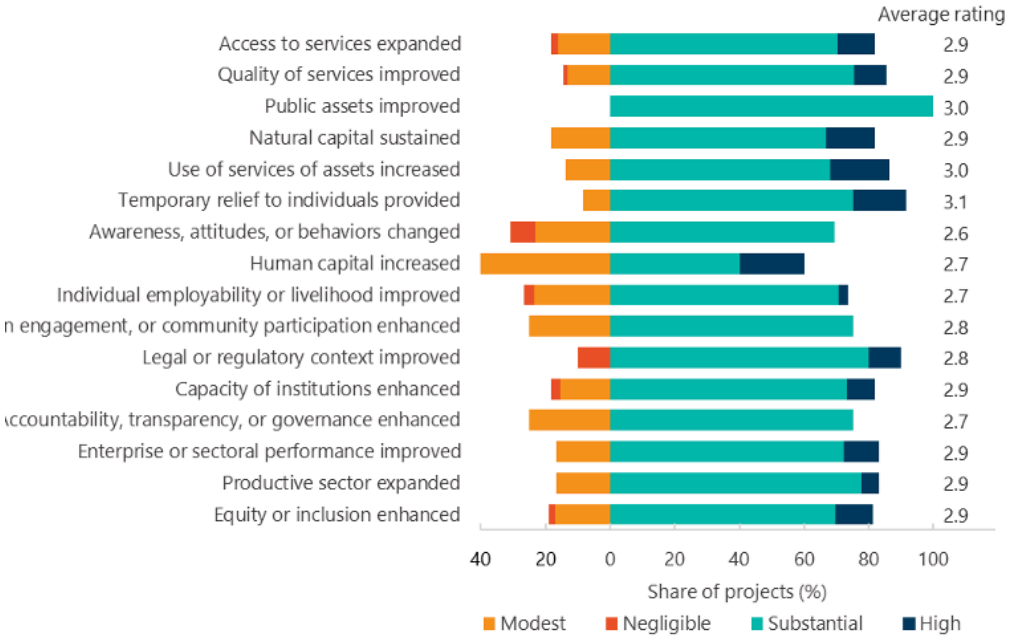
(continued)

Outcome Type	Share of Objectives (%)			Average Efficacy Rating					
	FY12–14	FY17–20(Q2)	FY20 (March)–22	FY12–14		FY17–20 (Q2)		FY20 (March)–22	
Citizen engagement enhanced	2	2	1	2.58	*α*	3.10	*α,β*	2.60	*β*
Legal or regulatory context improved	5	5	2	2.44	*γ*	2.46		3.00	*γ*
Capacity of institutions enhanced	37	40	33	2.43	*α,γ*	2.72	*α*	2.71	*γ*
Accountability or transparency enhanced	8	10	4	2.36	*α*	2.77	*α*	2.67	
Enterprise or sectoral performance improved	8	8	4	2.66		2.71		2.87	
Productive sector expanded	3	2	4	2.50		2.82		2.81	
Equity or inclusion enhanced	7	9	1	2.54	*α*	2.92	*α*	2.73	

Source: Independent Evaluation Group.

Note: FY12–14 and FY17–20(Q2) include only a sample of projects, representing 29 percent and 31 percent of the population, respectively. FY = fiscal year; Q2 = second quarter.
α. Statistically significant difference FY12–14 vs FY17–20(Q2).
γ. Statistically significant difference FY12–14 vs FY20 (March)–22.
β. Statistically significant difference FY17–20(Q2) vs FY20 (March)–22

Figure C.10. Overall Project Efficacy According to Outcome Type in the *Results and Performance of the World Bank Group 2023* Cohort

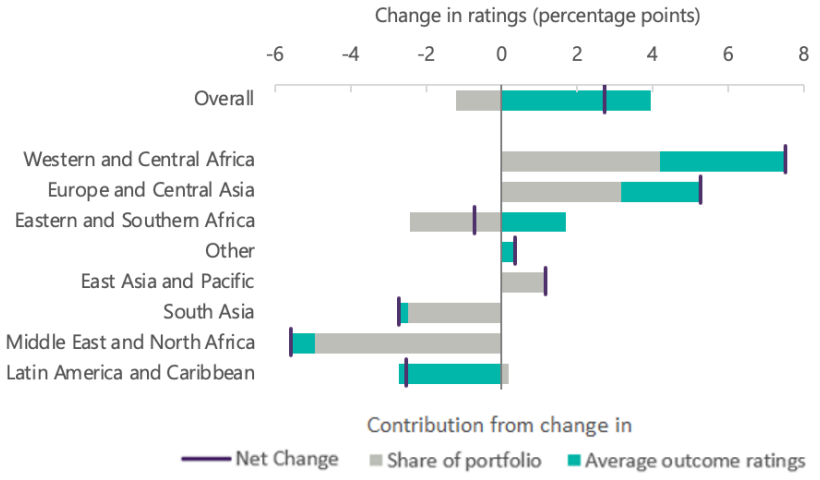


Source: Independent Evaluation Group.

Note: Efficacy is rated on a 4-point scale.

The improvement in project performance ratings between the prepandemic and *RAP 2023* cohorts is not attributed to a systematic difference in the composition of the portfolio. The decomposition analysis shows that the primary factor contributing to the overall increase in performance ratings is not portfolio changes but rating increases within various subgroups (including Global Practice, Region, project size, country income level, lending group, and fragile and conflict-affected situation status; figures C.11-C15).

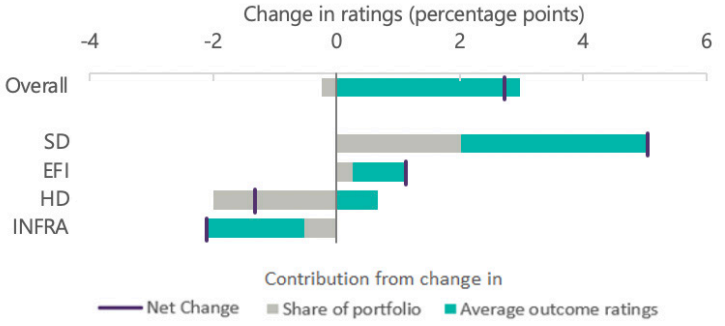
Figure C.11. Decomposition of Shift in Outcome Ratings between Prepandemic and *RAP 2023* Cohorts According to Region



Source: Independent Evaluation Group.

Note: RAP = Results and Performance of the World Bank Group.

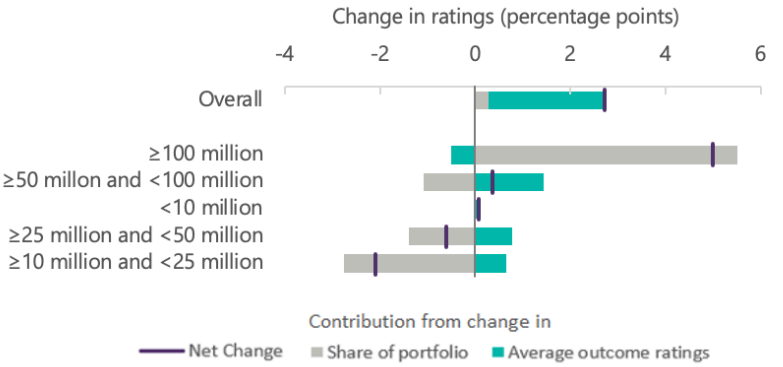
Figure C.12. Decomposition of Shift in Outcome Ratings between Prepandemic and *RAP 2023* Cohorts According to Practice Group



Source: Independent Evaluation Group.

Note: EFI= Equitable Growth, Finance and Institutions; HD = Human Development; INFRA = Infrastructure; *RAP = Results and Performance of the World Bank Group*; SD = Sustainable Development.

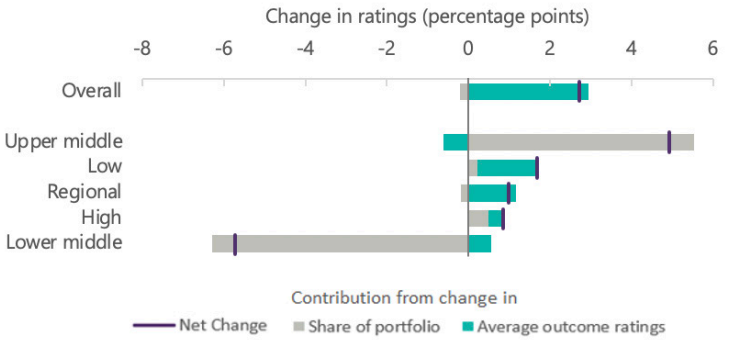
Figure C.13. Decomposition of Shift in Outcome Ratings between Prepandemic and *RAP 2023* Cohorts According to Project Size



Source: Independent Evaluation Group.

Note: RAP = Results and Performance of the World Bank Group.

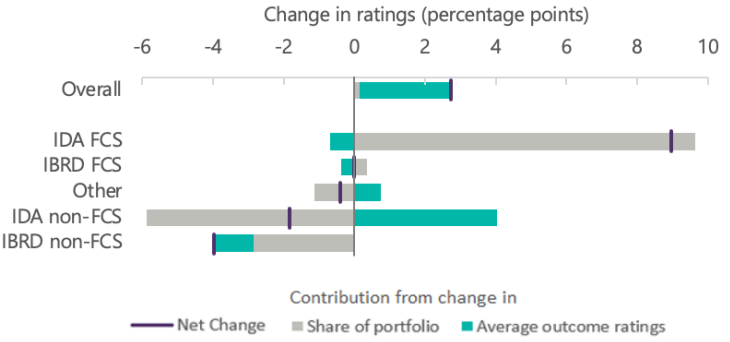
Figure C.14. Decomposition of Shift in Outcome Ratings between Prepandemic and *RAP 2023* Cohorts According to Country Income Group



Source: Independent Evaluation Group.

Note: RAP = Results and Performance of the World Bank Group.

Figure C.15. Decomposition of Shift in Outcome Ratings between Prepandemic and *RAP 2023* Cohorts According to Country Lending Group and FCS Status



Source: Independent Evaluation Group.

Note: FCS = fragile and conflict-affected situation; IBRD = International Bank for Reconstruction and Development; IDA = International Development Association; *RAP* = *Results and Performance of the World Bank Group*.

Efficacy, Monitoring and Evaluation Quality, and Adequacy of Results Frameworks

Efficacy and monitoring and evaluation quality ratings. This *RAP* found that, as in previous *RAP* reports, the quality of monitoring and evaluation (M&E) plays a crucial role in demonstrating the achievement of projects' intended development outcomes, as measured by efficacy ratings. There is a strong and positive correlation, backed by statistical significance, between efficacy and M&E quality ratings.[2] This result is to be expected, as the efficacy ratings rely heavily on the robustness of the evidence regarding the attainment of project objectives. A well-established and rigorous M&E framework is instrumental in collecting, analyzing, and presenting data that substantiates the evaluation of project effectiveness and the measurement of intended outcomes. Projects with higher M&E quality ratings are more likely to demonstrate a strong alignment between their stated objectives and the evidence of their accomplishment, resulting in higher efficacy ratings (table C.7).

Table C.7. Overall Efficacy and Monitoring and Evaluation Quality Ratings (percentage of projects)

	Overall Efficacy			
M&E Quality	Negligible	Modest	Substantial	High
Negligible	0.4	1.4	0.0	0.0
Modest	1.4	13.0	23.0	0.0
Substantial	0.0	1.4	45.0	7.0
High	0.0	0.0	4.0	4.0

Source: Independent Evaluation Group.

Note: M&E = monitoring and evaluation.

Nevertheless, cross-tabulations shows that approximately 23 percent of projects in the *RAP 2023* cohort (64 projects) were rated as having substantial efficacy but modest M&E quality ratings. Fifty-three percent of projects belong to the Sustainable Development Practice Group and 36 percent to the Infrastructure Practice Group (table C.8).

Table C.8. Distribution of Projects with Substantial Efficacy and Modest Monitoring and Evaluation Quality According to Practice Group

Practice Group	Projects (no.)	Share of Projects (%)
Equitable Growth, Finance, and Institutions	4	6
Human Development	3	5
Infrastructure	23	36
Sustainable Development	34	53
Total	64	100

Source: Independent Evaluation Group.

An in-depth review of a sample of 54 Independent Evaluation Group Implementation Completion and Results Report Reviews (ICRRs) was conducted to identify shortcomings in M&E practices in the M&E section as

well as additional evidence provided in the efficacy section. The detailed analysis of M&E revealed that the primary factor contributing to modest M&E quality ratings was deficiencies in M&E design, accounting for 93 percent of the cases examined. This indicates that the initial planning and structure of M&E systems were inadequate. Weak implementation of M&E practices, instead, was reported in 57 percent of projects and shortcomings in the use of M&E were identified in 19 percent of the projects (table C.9).

Table C.9. Shortcomings in Modest Monitoring and Evaluation Quality Ratings According to Practice Group (percentage)

Practice Group	Shortcomings in Design	Shortcomings in Implementation	Shortcoming in Use
Equitable Growth, Finance and Institutions	4	10	10
Human Development	2	10	0
Infrastructure	46	35	40
Sustainable Development	48	45	50
All Practice Groups	93	57	19

Source: Independent Evaluation Group.

In particular, the shortcomings in M&E design were related to (i) inadequate selection of indicators (for example, the outcome indicators were not comprehensive enough to capture the achievements of the objective, the indicators were output oriented rather than outcome oriented, and indicators did not show a strong link along a causal chain between outputs and outcomes); (ii) lack of methodology for collecting data; (iii) unrealistic targets; and (iv) attribution issues. Shortcomings in M&E implementation were related to (i) failure to rectify design shortcomings, (ii) delays in operationalizing M&E, (iii) inadequate M&E capacity and insufficient funds to support M&E activities, and (iv) weaknesses in data collection and data quality.

Despite modest M&E quality, efficacy was rated substantial for the following reasons:

1. Rectification of design shortcomings. In approximately half of the projects, efforts were made during the implementation phase to address design shortcomings. This involved refining the methodology, revising indicators, or adjusting targets through project restructuring. These revisions helped provide evidence to rate projects' efficacy as substantial.

2. Relying on output indicators. In cases in which outcome indicators were weak or lacking, substantial efficacy ratings were determined based on output indicators. These indicators assess whether project activities were completed and, in turn, whether these activities would plausibly contribute to achieving the desired outcomes as outlined in the project's theory of change. The substantial efficacy ratings were also based on additional information provided in the Implementation Completion and Results Report (ICR).

    - Evidence beyond the formal results framework. In 33 percent of the projects, the ICR included supporting evidence that extended beyond the formal results framework. This supplementary information reinforced the case for substantial efficacy.

    - Impact evaluations and beneficiary surveys. Impact evaluations or beneficiary surveys were conducted in 16 percent of the projects. These rigorous evaluations provided valuable evidence to support the substantial efficacy rating.

    - Additional evidence provided by the project team. In 12 percent of the projects, the project team offered supplementary evidence to substantiate the substantial efficacy rating, enhancing the overall assessment.

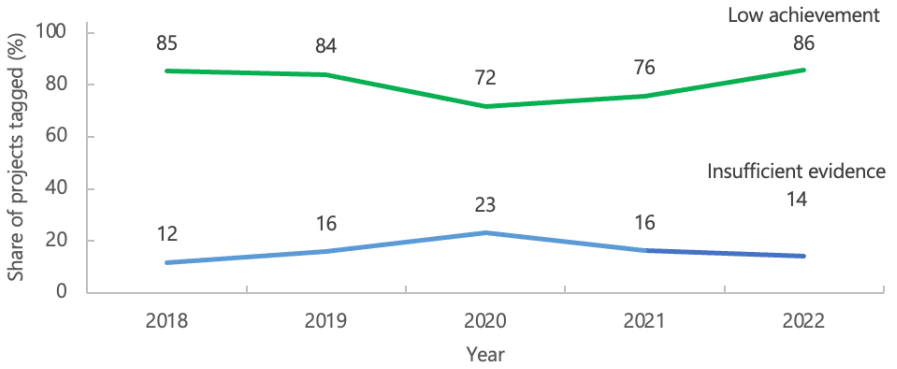
By considering these factors and leveraging various sources of evidence, including revised design, output indicators, additional information, impact evaluations, beneficiary surveys, and project team contributions, substantial efficacy ratings were determined, despite the modest quality of M&E.

Reason for low efficacy. An innovative aspect of this *RAP* is the first-time use of the ICRR data to identify the reasons underlying the efficacy ratings. This helped determine whether lower efficacy ratings reflected evidence of low achievement (that is, failed to achieve well-defined target indicators) or

resulted from insufficient evidence (that is, inadequate indicators and weak plausible attribution).

The justification for modest or negligible efficacy shows consistent findings with the improvements in M&E. In the *RAP 2023* cohort, 18 percent of projects had modest or negligible efficacy ratings. Low efficacy was mainly related to "evidence of low achievement" rather than "insufficient evidence." Moreover, since the introduction of the reason for low efficacy in the ICRR system in 2017, there have been no statistically significant changes in the share of unsuccessful projects for either reason (figure C.16).

Figure C.16. Reason for Projects with Overall Efficacy Score of Less than 3



Source: Independent Evaluation Group.

Efficacy ratings and validity of results frameworks. This *RAP*'s indicator mapping assessment shows that well-aligned indicators contributed to improved efficacy ratings (table C.10). Our analysis found a strong alignment between indicators and outcome types, with 97 percent of objectives having indicators of the same outcome type, and moreover, objectives with a larger share of indicators matching the objective outcome type pursued tend to have statistically significantly higher objective efficacy ratings. The adequacy of projects' results frameworks matters for both project and objective-level efficacy. Projects and objectives with a larger share of fully adequate indicators tend to have statistically significantly better efficacy ratings. However, there are no statistically significant associations between the indicator level and efficacy. One explanation is related to the above argument about the rating evaluation methodology, in which objectives lacking

outcome-level indicators or having weak indicators may still yield a substantial efficacy rating—provided that other lower-level indicators demonstrate completion of project activities and that these activities would plausibly contribute to achievement of outcomes as discussed under the theory of change and with triangulation of evidence. Another explanation is related to the nature of the intended development objective. Development objectives that envisage a change of intended intermediate outcomes do not need high outcome indicators to demonstrate achievement, which is, for example, the case of development objectives aimed to expand access to services.

Among the 16 outcome types, the World Bank has been more successful in achieving its goal of expanding access to services than any other outcome types (see table C.11). Expanding access to services was the intended development outcome with the highest efficacy rating (average efficacy of 3.06, which is a rating of substantial). Objectives pursuing this type of outcome also outperformed others in the adequacy of indicators. That said, many objectives related to expanding access to services were stated as lower-level results, thus not requiring high outcome indicators to measure and demonstrate achievements. On average, 74 percent of indicators measured outputs and intermediate outcomes, and 26 percent measured outcomes.

Table C.10. Validity of Indicators and Efficacy Ratings

	Rating			
Validity analysis	High	Substantial	Modest	Negligible
Project-level analysis	**Overall efficacy rating**			
Share of projects with outcome types matched by outcome types of indicators (%)	100	100	100	100
Indicator-level score				
All indicators (4-point scale)	1.82	1.81	1.82	1.80
PDO indicators (4-point scale)	2.42	2.33	2.24	2.14

(continued)

Validity analysis	Rating			
	High	Substantial	Modest	Negligible
Indicator adequacy score				
All indicators* (3-point scale)	2.31	2.30	2.20	2.32
Only with outcome-type-matched indicators* (3-point scale)	2.40	2.36	2.27	2.32
Objective-level analysis	**Objective efficacy rating**			
Share of objectives with outcome types matched by outcome types of indicators*	100	100	97	93
Indicator-level score				
All indicators (3-point scale)	1.84	1.82	1.80	1.90
PDO indicators (3-point scale)	2.36	2.34	2.24	2.23
Indicator adequacy score				
All indicators*	2.33	2.33	2.22	2.23
Only with outcome-type-matched indicators*	2.40	2.37	2.30	2.20
Share of fully adequate indicators				
All indicators* (%)	34	32	23	23
Only with outcome-type-matched indicators* (%)	40	37	30	20

Source: Independent Evaluation Group.

Note: PDO = project development objective.
** $p \leq .05$.

Objectives aimed to improve the quality of services also had substantial efficacy (average rating of 2.95) and high adequacy of indicators. Indicators measuring quality of services were found at all four levels in line with the specific dimensions of quality that the project focused on. For instance, quality of services objectives had on average 15 percent of indicators at the output level measuring improvements in structural quality, such as

rehabilitating or upgrading infrastructure and training service providers. Meanwhile, they had 48 percent of outcome-level indicators measuring, for example, time savings and user satisfaction with services provided, and only 5 percent of high outcome–level indicators measuring, for example, fatality rates.

Table C.11. Individual Objective Efficacy, Level, and Adequacy of Indicators

Top Three Development Outcomes	Average Individual Objective Efficacy Rating (4-point scale)	PDO Indicator-Level Score (4-point scale) and % of PDO Indicators According to Level	PDO Indicator Adequacy Score (3-point scale)
Access to services expanded	3.06 (*ß*)	2.20 (*α*) High outcome 0% Outcome 26% Intermediate outcome 68% Output 6%	2.78 (*ß*)
Quality of services improved	2.95	2.43 (*α,γ*) High outcome 5% Outcome 48% Intermediate outcome 32% Output 15%	2.70 (*γ*)
Capacity of institutions to perform institutional functions enhanced	2.76 (*ß*)	2.17 (*γ*) High outcome 0% Outcome 33% Intermediate outcome 50% Output 17%	2.46 (*ß,γ*)

Source: Independent Evaluation Group.

Note: Statistically significant at least 0.05 based on t-test and Mann-Whitney *U* test. *α* = statistically significant difference between access to services expanded and quality of services improved; *ß* = statistically significant difference between access to services expanded and capacity of institutions to perform institutional functions enhanced; *γ* = statistically significant between quality of services improved and capacity of institutions to perform institutional functions enhanced.

Enhancing the capacity of institutions to perform remains a particularly challenging outcome to achieve. Objectives targeting this outcome type received statistically significantly lower efficacy ratings than all other development outcomes (average rating of 2.76). One factor that contributes to the lower efficacy rating of this outcome type is the lower adequacy of the

results framework. Consistent with the findings of *RAP 2021*, this *RAP* found that the attainment of these development outcomes was measured predominantly by intermediate outcome or lower-level indicators (67 percent).

[1] Because projects can pursue multiple development outcomes, the proportions do not total 100 percent.

[2] Three types of correlation analyses were conducted to examine the relationship between efficacy and M&E quality ratings: Pearson correlation, Spearman's rank correlation, and Kendall's tau correlation. The coefficients obtained were all above 0.5, indicating a strong positive correlation. Additionally, the *p* values associated with these correlations were below 0.001, suggesting that the observed correlations are statistically significant.

Appendix D. Analysis of Factors Affecting World Bank Project Implementation and Performance

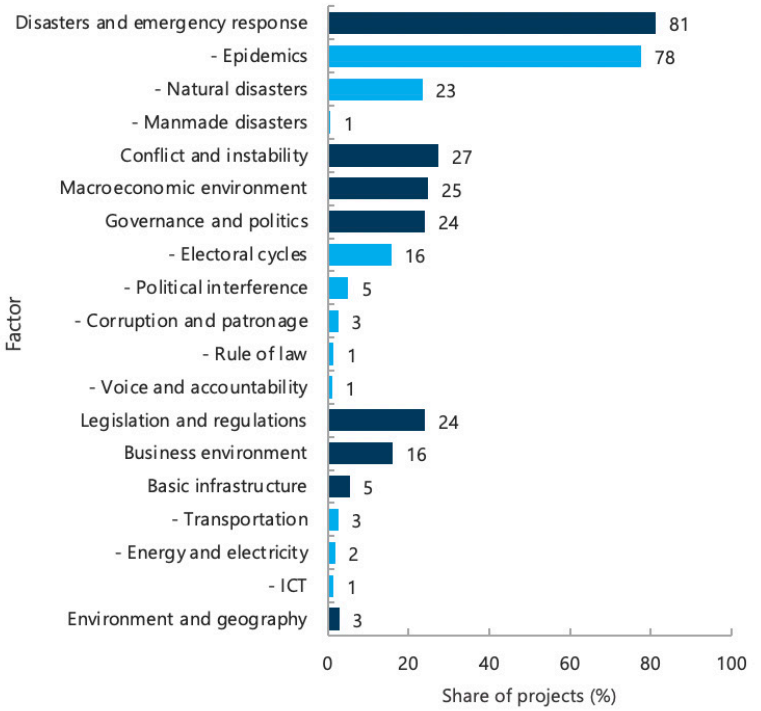
This appendix presents an in-depth analysis of the factors that affected the implementation and performance of projects within the *Results and Performance of the World Bank Group* (*RAP*) 2023 cohort during the COVID-19 pandemic, revealing patterns across projects and country characteristics. Furthermore, by comparing implementation factors before and after the pandemic, the report assesses whether there were differences in the types of enablers or challenges identified. This *RAP* also examines the correlation between these identified implementation factors and project performance ratings. Additionally, the analysis of project restructuring patterns provides insights into project performance ratings. (See appendix A for a detailed description of the methodology.)

Factors Affecting Project Implementation

Almost all projects in the *RAP 2023* cohort evaluated during fiscal years (FY)20–22 reported factors affecting implementation in all factor clusters. Specifically, 93 percent (255 projects) highlighted context-related factors, 96 percent (263 projects) include stakeholder-related factors, and 97 percent (265 projects) experienced project-related factors (figures D.1 and D.2).

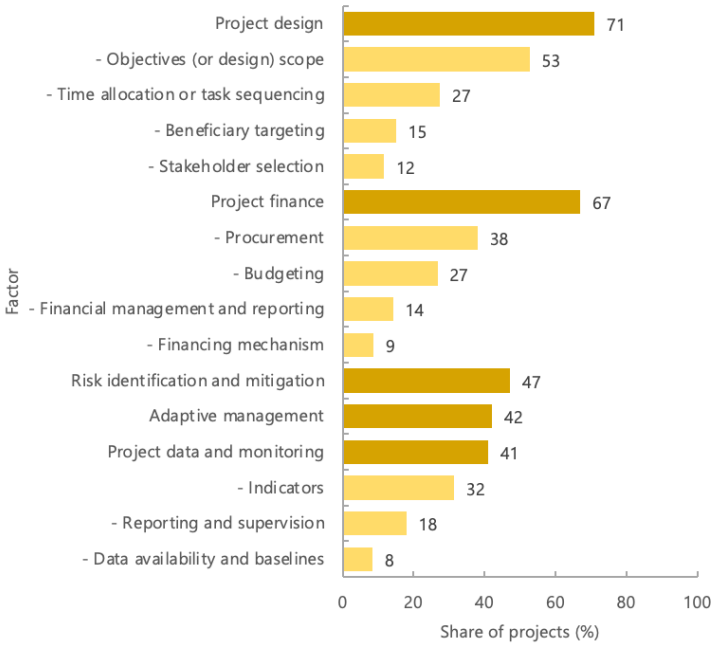
Figure D.1. Distribution of Factors Affecting Implementation in *RAP 2023* Projects

a. Context



b. Stakeholders

c. Project

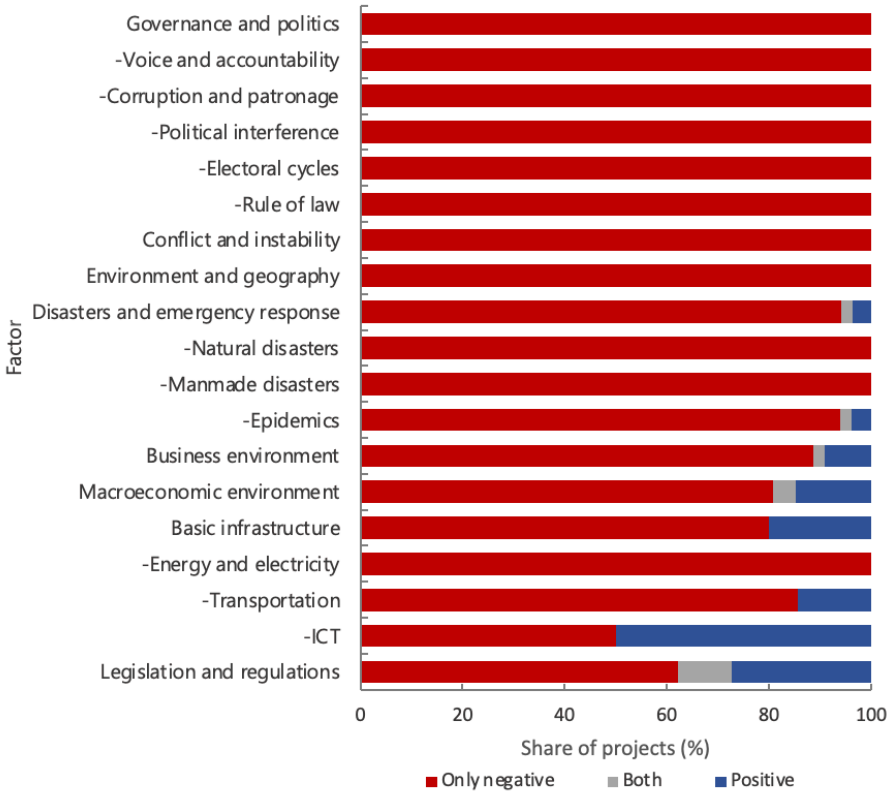


Source: Independent Evaluation Group.

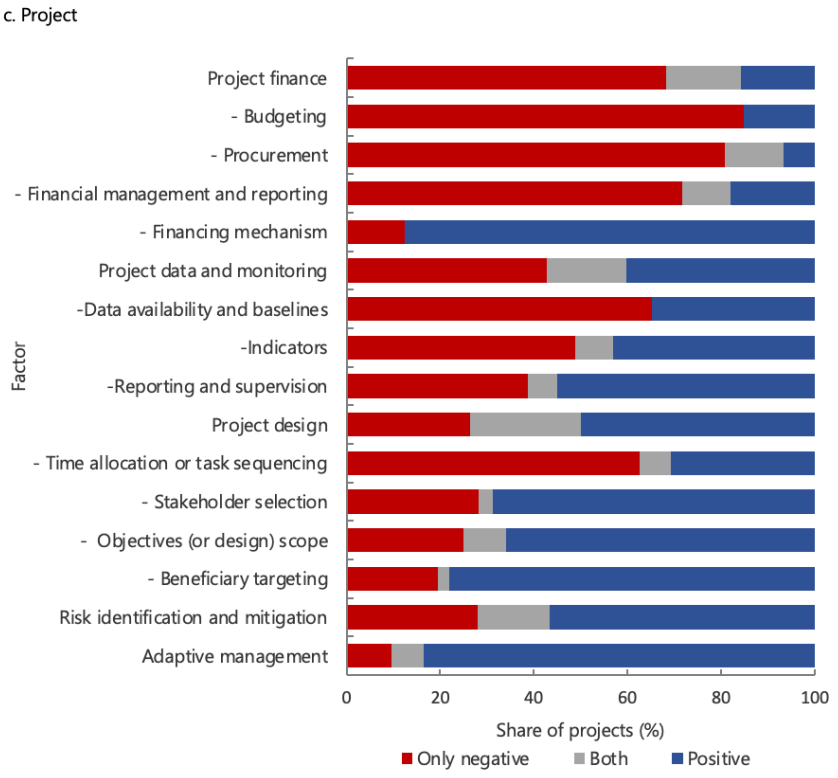
Note: ICT = information and communication technology.

Figure D.2. Sentiment of Factors Affecting Implementation in *RAP 2023* Projects

a. Context



b. Stakeholders



Source: Independent Evaluation Group.

Note: ICT = information and communication technology.

Context cluster. Despite projects' limited exposure to the pandemic (see box A.1), epidemics was the most salient contextual factor negatively affecting project implementation, reported by 212 projects (78 percent). Notably, although investment projects have experienced other outbreaks such as Ebola, cholera, and measles, the COVID-19 pandemic was by far the most frequently mentioned. Qualitative review of the epidemics subcategory revealed that lockdowns and mobility restrictions, economic downturn, disruptions in services and public institutions operations, implementation delays (particularly related to supply chain shortages and logistical challenges, and difficulty of in-person activities), and reallocation of project funds were specific ways that the pandemic affected project implementation (see box 2.1 in chapter 2). Other factors within the context cluster (such as natural disasters, conflict and instability, macroeconomic environment, governance and politics, legislation and regulations, business environment) also negatively affected

project implementation. However, these factors were only raised by a small portion of projects, accounting for less than 27 percent.

Stakeholder cluster. Approximately 74 percent of projects in the *RAP 2023* cohort reported implementation factors related to human resources and organizational capacity. Most projects (65 percent) identified challenges related to lack of technical capacity of implementing agencies, and 26 percent reported challenges resulting from frequent staff turnovers. Both factors mostly hindered implementation as they were reported with a negative sentiment.[1] Conversely, although factors related to skill transfer were less frequently mentioned (15 percent), they were reported as having facilitated implementation (positive sentiment). Coordination and engagement, as well as commitment and leadership, were identified as significant factors affecting implementation in 73 percent and 64 percent of projects, respectively. The sentiment associated with these factors was mixed, with approximately the same share of projects reporting them as either facilitators of or barriers to implementation.

Project cluster. Among project-related factors, project design issues were the most frequently reported in Implementation Completion and Results Reports (71 percent), with 35 percent of projects highlighting that the scope of the objectives or overall project design had predominantly facilitated implementation. Conversely, 27 percent of projects reported issues related to the time allocated or the sequencing of tasks required for project implementation, which were mostly expressed in a negative sentiment. Overall, project finance-related factors were highlighted by 67 percent of the projects. Among these, procurement-related factors were reported as affecting implementation by 38 percent of the projects, mostly hindering implementation (negative sentiment). Around 27 percent of projects faced implementation challenges related to funding gaps caused by inadequate budget provisions or delays in counterpart funding. Implementation factors related to project data and monitoring were reported by 41 percent of projects. Among these, indicators, availability of data, and reporting and supervision issues were mostly considered positive factors for implementation. Finally, ex ante risk identification and mitigation, and adaptations to unforeseen circumstances were highlighted by 47 and 42 percent of projects, respectively. Both factors were predominantly reported by project teams as facilitators of project

implementation. However, inadequate identification and mitigation of organizational capacity risks emerged as a challenge in project implementation. Interestingly, among the 56 projects that acknowledged the failure to adequately identify risks and implement effective mitigation measures, it was evident that the weak capacity of implementing agencies emerged as the most predominant underlying risk among other types of risks, accounting for 38 percent of cases (table D.1). These projects commonly reported that the initial risk assessments conducted before project implementation were overly optimistic given the complexity of the project. Consequently, the proposed mitigation measures proved insufficient, leading to delays in project implementation. Moreover, a significant majority of these projects, 15 out of 21 projects, also encountered obstacles in project execution precisely attributed to low technical and organizational capacity, captured by the skilled human resources and organizational capacities subcategory (figure D.3). This highlights the need for a comprehensive risk assessment and robust mitigation strategies, especially in countries where capacity limitations exist.

Importantly, all these context, stakeholder, and project-related factors, except for epidemics, cannot be attributed solely to the COVID-19 crisis. First, previous studies have also identified these factors in the past, indicating that these challenges are not unique to the COVID-19 pandemic.[2] Second, this *RAP*'s machine learning exercise, which expanded the analysis of factors to closed projects in FY18–20, revealed that similar factors were even more prevalently reported in the past than in projects exposed to the COVID-19 pandemic (see a comparative analysis of factors affecting implementation in the prepandemic and the *RAP 2023* cohort). Lastly, the implementation phase of the projects exposed to the pandemic spanned from 2003 onward, making it impossible to determine the specific timing of each factor, unlike epidemics, which has a specific starting point in March 2020.

Table D.1. Risks Insufficiently Identified and Mitigated

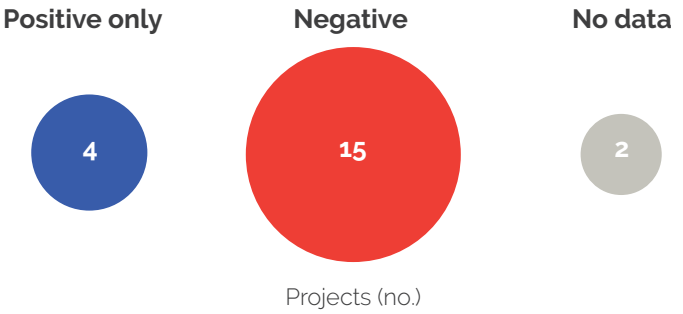
Risk Type	Share of Projects (*n* = 56; %)
Implementation capacity	38
Not specified	16
Political	13

(continued)

Risk Type	Share of Projects (*n* = 56; %)
Fiduciary	7
Environmental	5
Governance	5
Safeguards	5
Operational	4
Legislation	2
Economic	2
Stakeholders	2
Market response	2

Source: Independent Evaluation Group.

Figure D.3. Inadequate Risk Identification and Mitigation of Weak Institutional Capacity and Low Technical Capacity of Implementing Agencies (number of projects)



Source: Independent Evaluation Group.

Note: Negative = skilled human and organizational capacity factor reported as constraining implementation; positive = skilled human and organizational capacity factor reported as facilitating implementation; no data = skilled human and organizational capacity not reported.

Figures D.4–D.9 show the distribution of factors that hindered or facilitated implementation across Practice Groups and Global Practices, Regions, lending groups, country income level, and fragile and conflict-affected situation (FCS) status for the *RAP 2023* cohort.

Figure D.4. Distribution of Factors Affecting Implementation According to Practice Group

a. Frequency

	N	%	Equitable Growth, Finance and Institutions	Human Development	Infrastructure	Sustainable Development
Coordination and engagement	198	73	79	73	62	76
Commitment and leadership	176	64	82	76	53	59
Skilled human resources and organizational capacity	177	65	79	73	62	58
Skill transfer	40	15	12	34	17	5
Staff turnover	72	26	29	27	25	26
Legislation and regulations	66	24	15	24	18	30
Voice and accountability	3	1	0	2	0	2
Corruption and patronage	7	3	6	3	2	2
Political interference	14	5	3	10	3	4
Electoral cycles	43	16	24	15	12	16
Rule of law	4	1	0	0	2	3
Conflict and instability	75	27	35	29	22	28
Environment and geography	8	3	3	0	7	3
Information and communication technology	4	1	6	2	0	1
Energy and electricity	5	2	3	0	7	0
Transportation	7	3	3	5	0	3
Natural disasters	64	23	12	15	12	37
Manmade disasters	2	1	0	0	2	1
Epidemics	212	78	74	76	82	78
Business environment	44	16	15	3	22	20
Macroeconomic environment	68	25	32	25	17	27
Objectives (or design) scope	144	53	59	64	53	45
Time allocation or task sequencing	75	27	41	29	22	26
Stakeholder selection	32	12	18	12	8	12
Beneficiary targeting	41	15	6	41	0	13
Procurement	104	38	35	22	48	42
Financing mechanism	24	9	3	24	0	8
Budgeting	73	27	29	22	25	29
Financial management and reporting	39	14	18	19	8	14
Indicators	86	32	12	32	37	34
Data availability and baselines	23	8	0	7	12	10
Reporting and supervision	49	18	18	29	8	18
Adaptive management	115	42	29	59	25	46
Risk identification and mitigation	129	47	35	58	47	46

Subfactor tagged
0% 82%

(continued)

b. Sentiment

	N	Sentiment	Equitable Growth, Finance and Institutions	Human Development	Infrastructure	Sustainable Development
Coordination and engagement	198	0.1	0.1	0.3	-0.1	0.0
Commitment and leadership	176	0.0	0.3	0.1	-0.4	-0.1
Skilled human resources and organizational capacity	177	-0.4	-0.3	-0.4	-0.3	-0.6
Skill transfer	40	0.8	1.0	0.8	0.8	0.7
Staff turnover	72	-0.6	-0.3	-0.8	-0.6	-0.6
Legislation and regulations	66	-0.3	0.0	-0.3	-0.2	-0.5
Voice and accountability	3	-1.0		-1.0		-1.0
Corruption and patronage	7	-1.0	-1.0	-1.0	-1.0	-1.0
Political interference	14	-1.0	-1.0	-1.0	-1.0	-1.0
Electoral cycles	43	-1.0	-1.0	-1.0	-1.0	-1.0
Rule of law	4	-1.0			-1.0	-1.0
Conflict and instability	75	-1.0	-1.0	-1.0	-1.0	-1.0
Environment and geography	8	-1.0	-1.0		-1.0	-1.0
Information and communication technology	4	0.0	0.0	-1.0		1.0
Energy and electricity	5	-1.0	-1.0		-1.0	
Transportation	7	-0.7	-1.0	-1.0		-0.3
Natural disasters	64	-1.0	-1.0	-1.0	-1.0	-1.0
Manmade disasters	2	-1.0			-1.0	-1.0
Epidemics	212	-0.9	-0.9	-1.0	-0.7	-1.0
Business environment	44	-0.8	-0.2	0.0	-1.0	-0.9
Macroeconomic environment	68	-0.7	-1.0	-0.8	-0.4	-0.6
Objectives (or design) scope	144	0.4	-0.2	0.6	0.6	0.4
Time allocation or task sequencing	75	-0.3	-0.7	0.3	-0.5	-0.4
Stakeholder selection	32	0.4	0.0	1.0	1.0	0.1
Beneficiary targeting	41	0.6	1.0	0.8		0.3
Procurement	104	-0.7	-0.8	-0.6	-0.7	-0.8
Financing mechanism	24	0.8	1.0	0.9		0.6
Budgeting	73	-0.7	-0.2	-0.7	-0.7	-0.8
Financial management and reporting	39	-0.5	-0.2	-0.3	-0.4	-0.9
Indicators	86	-0.1	0.0	0.6	-0.3	-0.2
Data availability and baselines	23	-0.3		-0.5	0.4	-0.7
Reporting and supervision	49	0.2	0.0	0.4	-0.4	0.1
Adaptive management	115	0.7	0.6	0.9	0.9	0.6
Risk identification and mitigation	129	0.3	0.6	0.6	0.1	0.1

Subfactor sentiment

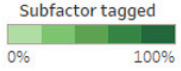
-1.0 1.0

Source: Independent Evaluation Group.

Figure D.5. Distribution of Factors Affecting Implementation According to Global Practice

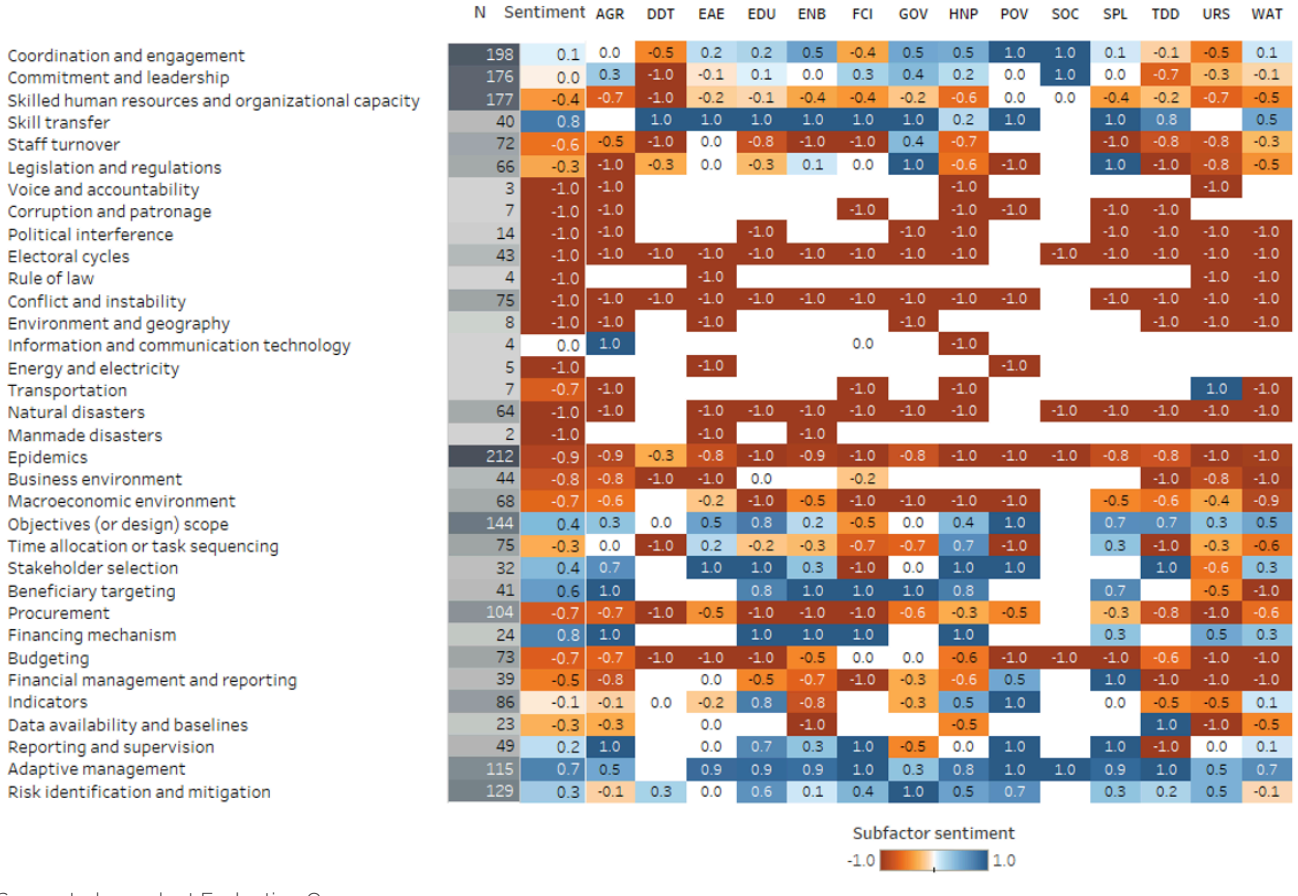
a. Frequency

	N	%	AGR	DDT	EAE	EDU	ENB	FCI	GOV	HNP	POV	SOC	SPL	TDD	URS	WAT
Coordination and engagement	198	73	74	67	60	59	89	70	91	90	100	100	69	62	68	77
Commitment and leadership	176	64	53	67	75	73	68	80	91	76	67	100	81	38	59	56
Skilled human resources and organizational capacity	177	65	53	50	60	77	58	70	91	76	100	50	63	65	68	51
Skill transfer	40	15	0	17	5	55	11	5	18	24	33	0	19	24	0	10
Staff turnover	72	26	21	17	20	23	26	25	45	43	0	0	13	29	27	28
Legislation and regulations	66	24	16	50	35	36	53	15	9	24	33	0	6	3	29	28
Voice and accountability	3	1	5	0	0	0	0	0	0	5	0	0	0	0	2	0
Corruption and patronage	7	3	11	0	0	0	0	5	0	5	33	0	6	3	0	0
Political interference	14	5	16	0	0	14	0	0	9	10	0	0	6	6	2	3
Electoral cycles	43	16	16	33	15	14	26	20	36	14	0	50	19	6	20	5
Rule of law	4	1	0	0	5	0	0	0	0	0	0	0	0	0	5	3
Conflict and instability	75	27	26	33	15	23	32	40	27	38	33	0	25	24	27	28
Environment and geography	8	3	5	0	5	0	0	0	9	0	0	0	0	9	2	3
Information and communication technology	4	1	5	0	0	0	0	10	0	5	0	0	0	0	0	0
Energy and electricity	5	2	0	0	20	0	0	0	0	0	33	0	0	0	0	0
Transportation	7	3	5	0	0	0	0	5	0	14	0	0	0	0	2	3
Natural disasters	64	23	42	0	10	9	16	10	18	24	0	50	13	15	39	41
Manmade disasters	2	1	0	0	5	0	5	0	0	0	0	0	0	0	0	0
Epidemics	212	78	95	100	80	86	84	65	91	71	67	100	69	79	71	72
Business environment	44	16	21	17	10	9	0	25	0	0	0	0	0	29	29	21
Macroeconomic environment	68	25	26	0	25	23	21	35	27	19	33	0	38	15	34	23
Objectives (or design) scope	144	53	16	50	65	64	26	55	73	71	33	0	56	47	51	64
Time allocation or task sequencing	75	27	21	33	25	23	21	35	55	29	33	0	38	18	22	36
Stakeholder selection	32	12	16	0	5	14	16	5	36	19	33	0	0	12	12	8
Beneficiary targeting	41	15	42	0	0	36	5	5	9	43	0	0	44	0	5	10
Procurement	104	38	37	50	55	27	26	25	45	14	67	0	25	44	41	54
Financing mechanism	24	9	5	0	0	45	5	5	0	5	0	0	19	0	10	8
Budgeting	73	27	32	33	10	9	42	20	36	43	67	50	13	32	27	23
Financial management and reporting	39	14	26	0	15	18	16	5	27	24	67	0	13	6	10	13
Indicators	86	32	37	17	55	41	26	0	27	38	33	0	13	29	27	46
Data availability and baselines	23	8	16	0	20	0	11	0	0	19	0	0	0	9	7	10
Reporting and supervision	49	18	5	0	15	32	16	5	36	38	33	0	13	6	15	28
Adaptive management	115	42	21	0	45	50	42	10	55	62	67	50	69	18	49	56
Risk identification and mitigation	129	47	37	67	55	77	53	35	18	52	100	0	38	38	49	46



(continued)

b. Sentiment



Source: Independent Evaluation Group.

Note: AGR = Agriculture and Food; DDT = Digital Development; EAE = Energy and Extractives; EDU = Education; ENB = Environment, Natural Resources and Blue Economy; FCI = Finance, Competitiveness, and Innovation; GOV = Governance; HNP = Health, Nutrition, and Population; POV = Poverty and Equity; SOC = Social Sustainability and Inclusion; SPL = Social Protection and Jobs; TDD = Transport; URS = Urban, Disaster Risk Management, Resilience,, and Land; WAT = Water.

Figure D.6. Distribution of Factors Affecting Implementation According to Region (percentage of projects)

a. Frequency

	N	%	Eastern and Southern Africa	Western and Central Africa	East Asia and Pacific	Europe and Central Asia	Latin America and Caribbean	Middle East and North Africa	South Asia	Other
Coordination and engagement	198	73	70	76	76	68	67	88	74	0
Commitment and leadership	176	64	70	68	60	68	77	38	51	0
Skilled human resources and organizational capacity	177	65	75	68	62	51	56	25	85	0
Skill transfer	40	15	15	20	12	11	5	13	23	0
Staff turnover	72	26	30	25	28	14	38	13	26	0
Legislation and regulations	66	24	25	14	26	41	28	25	18	0
Voice and accountability	3	1	5	2	0	0	0	0	0	0
Corruption and patronage	7	3	5	3	0	0	3	13	3	0
Political interference	14	5	10	3	2	5	3	25	5	0
Electoral cycles	43	16	15	22	0	19	21	0	23	0
Rule of law	4	1	5	2	0	0	0	0	3	0
Conflict and instability	75	27	30	46	2	11	26	100	33	0
Environment and geography	8	3	3	2	4	3	3	0	5	0
Information and communication technology	4	1	3	0	4	0	3	0	0	0
Energy and electricity	5	2	3	0	2	3	3	0	3	0
Transportation	7	3	5	5	4	0	0	0	0	0
Natural disasters	64	23	33	15	18	11	36	13	36	0
Manmade disasters	2	1	0	0	0	0	3	0	3	0
Epidemics	212	78	78	81	74	70	85	88	77	0
Business environment	44	16	18	14	8	22	15	25	23	0
Macroeconomic environment	68	25	40	22	8	19	33	63	26	0
Objectives (or design) scope	144	53	55	49	56	54	62	38	44	100
Time allocation or task sequencing	75	27	38	25	28	30	26	13	23	0
Stakeholder selection	32	12	10	15	8	16	13	13	8	0
Beneficiary targeting	41	15	15	24	10	11	15	13	13	0
Procurement	104	38	40	34	28	49	33	38	51	0
Financing mechanism	24	9	5	14	6	5	10	0	13	0
Budgeting	73	27	33	41	26	16	26	13	15	0
Financial management and reporting	39	14	15	19	10	11	13	13	18	0
Indicators	86	32	25	27	42	27	38	38	28	0
Data availability and baselines	23	8	3	5	14	5	13	0	13	0
Reporting and supervision	49	18	15	27	22	11	13	13	15	0
Adaptive management	115	42	35	53	34	38	41	75	44	0
Risk identification and mitigation	129	47	50	44	60	35	41	50	49	100

Subfactor tagged

0% 100%

(continued)

b. Sentiment

	N	Sentiment	Eastern and Southern Africa	Western and Central Africa	East Asia and Pacific	Europe and Central Asia	Latin America and Caribbean	Middle East and North Africa	South Asia	Other
Coordination and engagement	198	0.1	0.0	0.5	0.0	-0.2	-0.1	0.0	0.0	
Commitment and leadership	176	0.0	0.0	0.2	0.2	-0.2	-0.3	-1.0	-0.1	
Skilled human resources and organizational capacity	177	-0.4	-0.5	-0.6	-0.3	-0.2	-0.4	0.0	-0.5	
Skill transfer	40	0.8	0.3	1.0	0.7	1.0	1.0	1.0	0.8	
Staff turnover	72	-0.6	-0.9	-0.5	-0.3	-0.2	-0.6	-1.0	-1.0	
Legislation and regulations	66	-0.3	-0.4	-0.5	-0.3	-0.1	-0.5	-1.0	-0.1	
Voice and accountability	3	-1.0	-1.0	-1.0						
Corruption and patronage	7	-1.0	-1.0	-1.0			-1.0	-1.0	-1.0	
Political interference	14	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	
Electoral cycles	43	-1.0	-1.0	-1.0		-1.0	-1.0		-1.0	
Rule of law	4	-1.0	-1.0	-1.0					-1.0	
Conflict and instability	75	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	
Environment and geography	8	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0		-1.0	
Information and communication technology	4	0.0	-1.0		0.0		1.0			
Energy and electricity	5	-1.0	-1.0		-1.0	-1.0	-1.0		-1.0	
Transportation	7	-0.7	-1.0	-1.0	0.0					
Natural disasters	64	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	
Manmade disasters	2	-1.0					-1.0		-1.0	
Epidemics	212	-0.9	-0.9	-1.0	-0.8	-0.8	-0.9	-0.4	-1.0	
Business environment	44	-0.8	-0.4	-0.9	-1.0	-1.0	-1.0	-1.0	-0.6	
Macroeconomic environment	68	-0.7	-0.7	-0.8	-0.5	-0.7	-0.8	-1.0	-0.1	
Objectives (or design) scope	144	0.4	0.5	0.4	0.5	0.7	0.0	0.7	0.2	1.0
Time allocation or task sequencing	75	-0.3	-0.7	-0.1	-0.4	-0.1	-0.5	1.0	-0.2	
Stakeholder selection	32	0.4	0.5	0.3	0.0	0.7	0.2	0.0	1.0	
Beneficiary targeting	41	0.6	0.0	0.7	1.0	0.5	0.7	1.0	0.4	
Procurement	104	-0.7	-0.9	-0.8	-0.6	-0.7	-0.9	-1.0	-0.6	
Financing mechanism	24	0.8	0.0	1.0	1.0	0.0	0.5		1.0	
Budgeting	73	-0.7	-1.0	-0.8	-0.5	-0.3	-0.6	-1.0	-0.7	
Financial management and reporting	39	-0.5	-0.2	-0.5	-1.0	-1.0	-0.6	0.0	-0.3	
Indicators	86	-0.1	-0.2	0.4	0.0	0.1	-0.7	-0.3	0.0	
Data availability and baselines	23	-0.3	1.0	-0.3	0.4	-1.0	-0.6		-1.0	
Reporting and supervision	49	0.2	0.2	0.1	0.2	0.8	0.2	1.0	-0.3	
Adaptive management	115	0.7	0.8	0.7	0.9	0.9	0.6	0.8	0.5	
Risk identification and mitigation	129	0.3	0.1	0.7	0.4	0.7	-0.2	0.0	0.0	1.0

Subfactor sentiment

Source: Independent Evaluation Group.

Figure D.7. Distribution of Factors Affecting Implementation According to Lending Group

a. Frequency

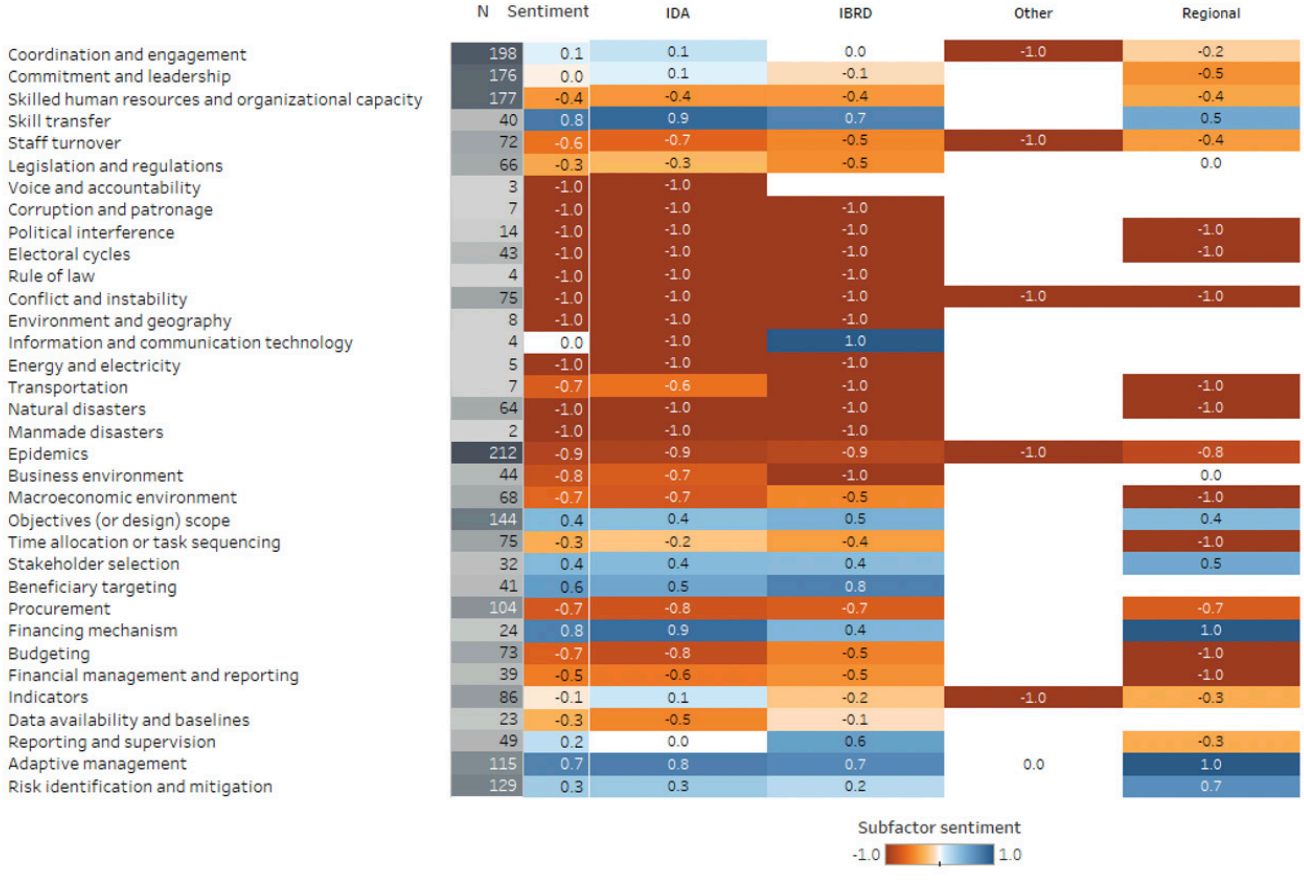
	N	%	IDA	IBRD	Other	Regional
Coordination and engagement	198	73	72	72	100	76
Commitment and leadership	176	64	66	64	0	65
Skilled human resources and organizational capacity	177	65	71	57	0	65
Skill transfer	40	15	14	14	0	24
Staff turnover	72	26	29	22	100	29
Legislation and regulations	66	24	20	29	0	29
Voice and accountability	3	1	2	0	0	0
Corruption and patronage	7	3	3	2	0	0
Political interference	14	5	6	4	0	6
Electoral cycles	43	16	17	15	0	12
Rule of law	4	1	1	2	0	0
Conflict and instability	75	27	39	13	100	24
Environment and geography	8	3	3	3	0	0
Information and communication technology	4	1	1	2	0	0
Energy and electricity	5	2	2	2	0	0
Transportation	7	3	3	1	0	6
Natural disasters	64	23	28	20	0	6
Manmade disasters	2	1	1	1	0	0
Epidemics	212	78	80	75	100	76
Business environment	44	16	17	15	0	12
Macroeconomic environment	68	25	27	24	0	18
Objectives (or design) scope	144	53	54	51	0	59
Time allocation or task sequencing	75	27	30	25	0	24
Stakeholder selection	32	12	10	12	0	24
Beneficiary targeting	41	15	20	11	0	0
Procurement	104	38	39	37	0	41
Financing mechanism	24	9	11	6	0	6
Budgeting	73	27	30	26	0	6
Financial management and reporting	39	14	19	10	0	6
Indicators	86	32	30	35	100	18
Data availability and baselines	23	8	8	10	0	0
Reporting and supervision	49	18	21	15	0	18
Adaptive management	115	42	46	39	100	24
Risk identification and mitigation	129	47	48	46	0	53

Subfactor tagged

0% 100%

(continued)

b. Sentiment



Source: Independent Evaluation Group.

Note: IDA = International Development Association; IBRD = International Bank for Reconstruction and Development.

Figure D.8. Distribution of Factors Affecting Implementation According to Country Income Group

a. Frequency

	N	%	High	Upper middle	Lower middle	Low	Regional
Coordination and engagement	198	73	100	67	77	70	76
Commitment and leadership	176	64	75	70	57	70	65
Skilled human resources and organizational capacity	177	65	75	51	72	70	65
Skill transfer	40	15	0	11	17	14	24
Staff turnover	72	26	25	24	30	23	29
Legislation and regulations	66	24	50	29	25	14	29
Voice and accountability	3	1	0	0	0	5	0
Corruption and patronage	7	3	0	2	2	5	0
Political interference	14	5	0	2	6	8	6
Electoral cycles	43	16	25	12	17	18	12
Rule of law	4	1	0	1	2	2	0
Conflict and instability	75	27	25	11	28	48	24
Environment and geography	8	3	0	2	5	2	0
Information and communication technology	4	1	0	1	2	2	0
Energy and electricity	5	2	0	2	1	3	0
Transportation	7	3	0	0	2	6	6
Natural disasters	64	23	0	17	28	30	6
Manmade disasters	2	1	0	1	1	0	0
Epidemics	212	78	50	76	80	79	76
Business environment	44	16	0	16	15	20	12
Macroeconomic environment	68	25	0	22	26	30	18
Objectives (or design) scope	144	53	50	50	54	53	59
Time allocation or task sequencing	75	27	50	24	26	33	24
Stakeholder selection	32	12	25	13	8	12	24
Beneficiary targeting	41	15	0	12	16	21	0
Procurement	104	38	25	34	32	53	41
Financing mechanism	24	9	0	5	15	5	6
Budgeting	73	27	25	26	31	27	6
Financial management and reporting	39	14	0	10	20	14	6
Indicators	86	32	0	35	34	29	18
Data availability and baselines	23	8	0	12	12	2	0
Reporting and supervision	49	18	0	17	20	17	18
Adaptive management	115	42	25	38	44	50	24
Risk identification and mitigation	129	47	25	45	50	45	53

Subfactor tagged

0% 100%

(continued)

b. Sentiment

	N	Sentiment	High	Upper middle	Lower middle	Low	Regional
Coordination and engagement	198	0.1	0.0	0.0	0.1	0.2	-0.2
Commitment and leadership	176	0.0	-0.7	-0.2	0.1	0.0	-0.5
Skilled human resources and organizational capacity	177	-0.4	-0.7	-0.3	-0.5	-0.3	-0.4
Skill transfer	40	0.8		0.8	0.9	0.8	0.5
Staff turnover	72	-0.6	1.0	-0.6	-0.6	-0.7	-0.4
Legislation and regulations	66	-0.3	0.0	-0.5	-0.5	0.1	0.0
Voice and accountability	3	-1.0				-1.0	
Corruption and patronage	7	-1.0		-1.0	-1.0	-1.0	
Political interference	14	-1.0		-1.0	-1.0	-1.0	-1.0
Electoral cycles	43	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
Rule of law	4	-1.0		-1.0	-1.0	-1.0	
Conflict and instability	75	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
Environment and geography	8	-1.0		-1.0	-1.0	-1.0	
Information and communication technology	4	0.0		1.0	0.0	-1.0	
Energy and electricity	5	-1.0		-1.0	-1.0	-1.0	
Transportation	7	-0.7			0.0	-1.0	-1.0
Natural disasters	64	-1.0		-1.0	-1.0	-1.0	-1.0
Manmade disasters	2	-1.0		-1.0	-1.0		
Epidemics	212	-0.9	-1.0	-0.9	-0.9	-0.9	-0.8
Business environment	44	-0.8		-1.0	-0.7	-0.8	0.0
Macroeconomic environment	68	-0.7		-0.8	-0.6	-0.7	-1.0
Objectives (or design) scope	144	0.4	0.0	0.4	0.5	0.3	0.4
Time allocation or task sequencing	75	-0.3	-1.0	-0.4	-0.3	-0.1	-1.0
Stakeholder selection	32	0.4	-1.0	0.5	0.5	0.4	0.5
Beneficiary targeting	41	0.6		0.8	0.6	0.4	
Procurement	104	-0.7	-1.0	-0.8	-0.6	-0.8	-0.7
Financing mechanism	24	0.8		0.0	0.9	1.0	1.0
Budgeting	73	-0.7	-1.0	-0.5	-0.7	-0.9	-1.0
Financial management and reporting	39	-0.5		-0.8	-0.5	-0.4	-1.0
Indicators	86	-0.1		-0.3	0.1	0.1	-0.3
Data availability and baselines	23	-0.3		0.0	-0.5	-1.0	
Reporting and supervision	49	0.2		0.4	0.2	-0.2	-0.3
Adaptive management	115	0.7	-1.0	0.8	0.7	0.7	1.0
Risk identification and mitigation	129	0.3	-1.0	0.4	0.0	0.5	0.7

Source: Independent Evaluation Group.

Figure D.9. Distribution of Factors Affecting Implementation According to Fragile and Conflict-Affected Situation Status

a. Frequency

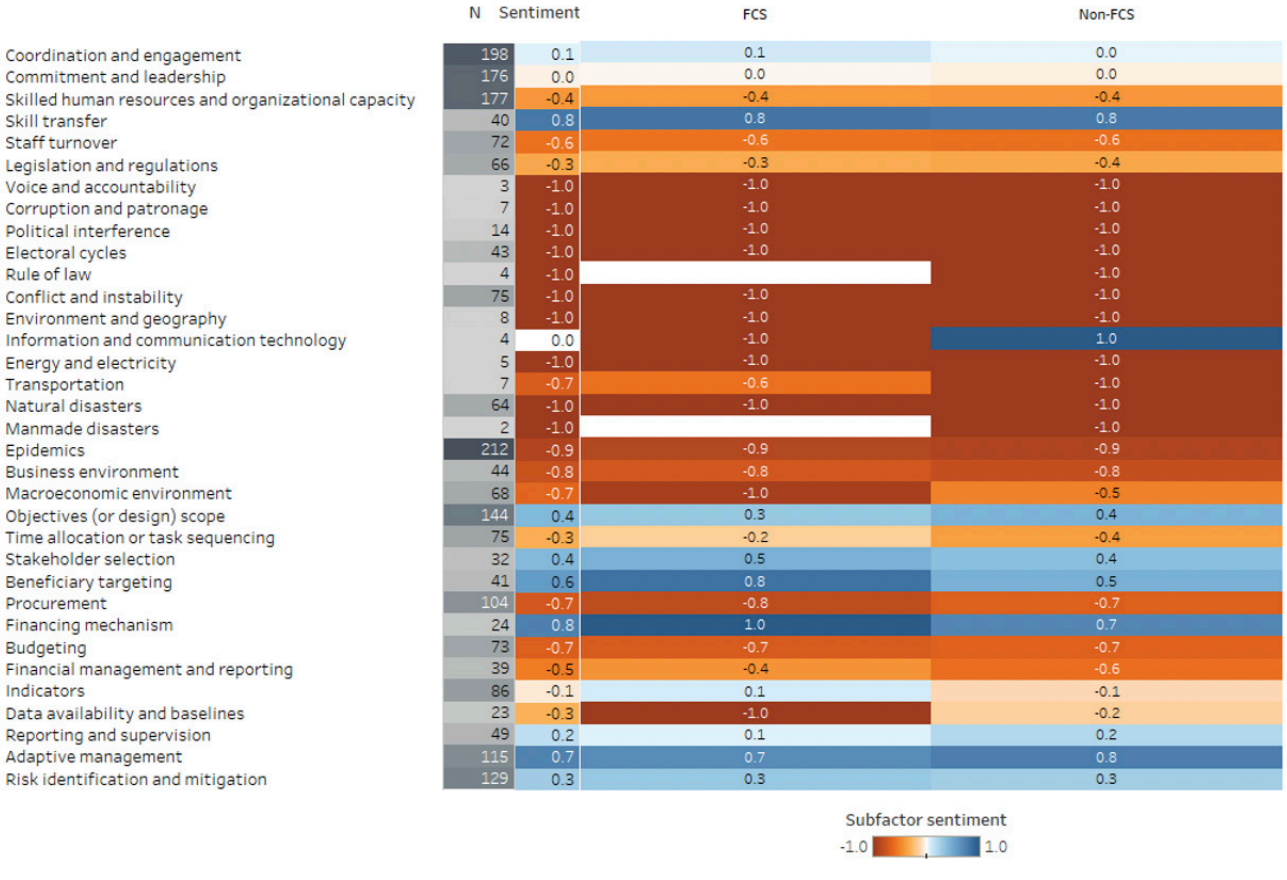
	N	%	FCS	Non-FCS
Coordination and engagement	198	73	70	73
Commitment and leadership	176	64	69	63
Skilled human resources and organizational capacity	177	65	70	63
Skill transfer	40	15	14	15
Staff turnover	72	26	31	24
Legislation and regulations	66	24	19	26
Voice and accountability	3	1	3	1
Corruption and patronage	7	3	8	1
Political interference	14	5	9	4
Electoral cycles	43	16	22	13
Rule of law	4	1	0	2
Conflict and instability	75	27	60	15
Environment and geography	8	3	1	4
Information and communication technology	4	1	3	1
Energy and electricity	5	2	3	2
Transportation	7	3	6	1
Natural disasters	64	23	23	23
Manmade disasters	2	1	0	1
Epidemics	212	78	79	77
Business environment	44	16	16	16
Macroeconomic environment	68	25	29	23
Objectives (or design) scope	144	53	53	53
Time allocation or task sequencing	75	27	31	26
Stakeholder selection	32	12	17	10
Beneficiary targeting	41	15	17	14
Procurement	104	38	40	37
Financing mechanism	24	9	4	11
Budgeting	73	27	29	26
Financial management and reporting	39	14	21	12
Indicators	86	32	31	32
Data availability and baselines	23	8	5	10
Reporting and supervision	49	18	21	17
Adaptive management	115	42	49	39
Risk identification and mitigation	129	47	49	46

Subfactor tagged

0% 79%

(continued)

b. Sentiment



Source: Independent Evaluation Group.

Note: FCS = fragile and conflict-affected situations.

Context-related factors consistently hindered project implementation, and this pattern holds across various project groups. Nonetheless, there are variations in the prevalence of these contextual factors. For instance, the disruptions caused by the COVID-19 pandemic, referred to as "epidemics," were more frequently reported in Latin America and the Caribbean and the Middle East and North Africa. Implementation challenges associated with natural disasters were more commonly reported in Sustainable Development projects than in other Practice Groups, particularly in Eastern and Southern Africa, Latin America and the Caribbean, and South Asia and in low-income and lower-middle-income countries. Factors related to conflict and instability were more prevalent in Western and Central Africa and International Development Association (IDA) and FCS countries. Destabilizing factors associated with the macroeconomic environment were more frequently reported in the Middle East and North Africa and Eastern and Southern Africa. Factors related to electoral cycles were prominent in the Equitable Growth, Finance, and Institutions sector and were also notable in Western and Central Africa, Latin America and the Caribbean, and South Asia and in FCS countries. Legislation and regulation factors were more prevalent in Europe and Central Asia.

Regarding stakeholder dynamics, challenges related to skilled human resources and organizational capacity were more prominent in Equitable Growth, Finance, and Institutions than in other Practice Groups. They were particularly notable in South Asia and Eastern and Southern Africa and in IDA and FCS countries. Meanwhile, coordination and engagement challenges were more prevalent in the Middle East and North Africa. Coordination and engagement had a positive impact on project implementation in the Equitable Growth, Finance, and Institutions and Human Development Practice Groups but a negative impact in the Infrastructure and Sustainable Development Practice Groups; Europe and Central Asia, Latin America and the Caribbean, and South Asia; IDA countries; and upper-middle-income countries. Similarly, commitment and leadership had a positive effect on implementation in the Equitable Growth, Finance, and Institutions and Human Development Practice Groups but a negative impact in the Infrastructure and Sustainable Development Practice Groups; Europe and Central Asia,

Latin America and the Caribbean, and South Asia Regions; IDA countries; and upper-middle-income countries.

Project-related factors generally followed a consistent pattern across project groups, although there were a few exceptions. Risk identification and mitigation factors were more frequently reported in projects in East Asia and Pacific, typically having positive effects on implementation. However, in Latin America and the Caribbean, risk identification and mitigation were typically reported as hindering project implementation. Adaptive management was more commonly reported in projects in the Human Development and Sustainable Development Practice Groups, particularly in the Middle East and North Africa. As anticipated, factors related to beneficiary targeting were more prevalent in projects in the Human Development Practice Group.

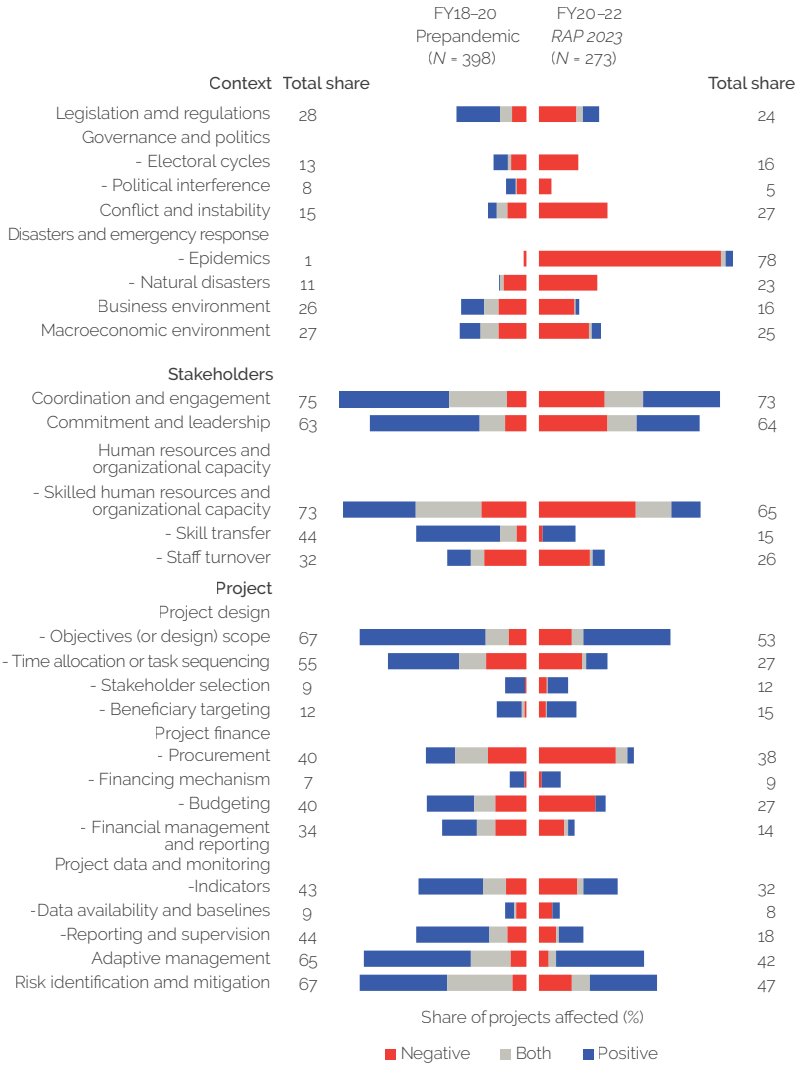
Factors Affecting Implementation of Projects in the Prepandemic and *RAP 2023* Cohorts: A Supervised Machine Learning Exercise

Comparing the salient factors affecting implementation during the COVID-19 pandemic is crucial to understanding whether they differed from factors in prepandemic times. This *RAP* used supervised machine learning techniques to expand the analysis of implementation factors in previous cohorts. (See appendix A for the methodology of supervised machine learning.)

The COVID-19 pandemic did not seem to increase the prevalence of factors affecting the implementation of projects in FY20–22. The results of the supervised machine learning model indicate that a similar or slightly larger share of projects reported implementation factors in the prepandemic periods, with the exception of three specific contextual factors (epidemics, conflict and instability, and natural disasters; figure D.10). However, a greater share of projects exposed to the COVID-19 pandemic reported that these factors hindered implementation than in the prepandemic cohort. Factors that on average were reported as facilitators of project implementation in the prepandemic period tended to have a more neutral effect on implementation in FY20–22 projects (mixed sentiment). Likewise, factors that were previously reported as neutral to implementation were highlighted in Implementation Completion and Results Reports as hindering

implementation (negative sentiment) of projects exposed to the COVID-19 pandemic. Only a few factors remained mostly positive in affecting implementation: adaptive management, risk identification and mitigation, financing mechanism, and skill transfer.

Figure D.10. Factors Affecting Project Implementation: Prepandemic versus *RAP 2023* Cohort



Source: Independent Evaluation Group.

Note: Negative = the identified factor was reported as a constraint to project implementation. Positive = the identified factor was reported as facilitating implementation. Both = at the project level, there were positive and negative factors in the same category. This is more prominent in categories that were not disaggregated, such as coordination and engagement. For example, the Implementation Completion and Results Report showed that there was a clear allocation of roles and responsibilities (positive), but the bureaucratic structure created challenges to project implementation (negative). *RAP = Results and Performance of the World Bank Group*.

Implementation Factors and Project Performance Ratings

Factors that affected project implementation in the *RAP 2023* cohort had a limited effect on project performance ratings. Only 28 percent of the factors identified show a significant association with project ratings (figure D.11; box D.1). The two most reported factors were skilled human resources and organizational capacity and risk identification and mitigation. There was a statistically significant difference in average ratings between projects that identified the implementation factor and those that did not, as determined using Student *t*-tests and Mann-Whitney *U*-tests, although the difference was small.

Figure D.11. Implementation Factors Statistically Significantly Associated with Project Performance Ratings

Source: Independent Evaluation Group.

Note: Differences in average ratings between projects that identified the implementation factor and those that did not were statistically significant, as determined by both t-test and Mann-Whitney U test. M&E = monitoring and evaluation; *RAP = Results and Performance of the World Bank Group*.

Box D.1. Implementation Factors Statistically Associated with Project Performance Ratings

Context-Related Factors

Governance and politics. Twenty-four percent of projects in the *Results and Performance of the World Bank Group 2023* cohort reported challenges related to governance and politics, for example, electoral cycles, limited accountability, corrupt practices, and political manipulation of projects, that were beyond the control of the World Bank. The impact of these challenges on project performance ratings was minor and, as expected, was not correlated with Bank performance ratings. That said, projects that reported challenges related to governance and politics had slightly lower outcome ratings (average 4.15) than those without such challenges (average 4.43). These findings are consistent with those by Ortega Nieto et al. (2022), who found that governance and politics have a negligible impact on project performance.

Natural disasters. About 23 percent of projects faced natural disasters during implementation. Natural hazards such as droughts, floods, cyclones, and earthquakes can significantly affect monitoring and evaluation (M&E) as well as supervision efforts because damage to the road infrastructure and emergency and priority reconstruction activities can make data collection and analysis difficult. These challenges resulted in a minor decrease in M&E quality (average 2.52) and quality of supervision ratings (average 4.39) than in projects that did not report challenges regarding natural disasters (rating averages 2.71 and 4.64, respectively).

Stakeholder-Related Factors

Skilled human resources and organizational capacity. The limited technical capacity of implementing agencies to supervise the quality of work can result in systemic delays in implementation and suboptimal outcomes, as indicated by 65 percent of projects reporting skilled human resources and organizational capacity as critical factors. Most of these projects indicated that lack of technical capacity was a major implementation constraint. Weak implementation capacity was particularly predominant in South Asia and Eastern and Southern Africa regions, in International Development Association countries, and in countries classified as fragile and conflict-affected situations (figures D.5, D.6, and D.8).

(continued)

Box D.1. Implementation Factors Statistically Associated with Project Performance Ratings (cont.)

Projects reporting this factor had an average outcome rating of 4.25 (in the range of moderately satisfactory), and those that did not had an average outcome rating of 4.55 (in the range of satisfactory). Previous studies, including those by Denizer et al. (2013) and Ortega Nieto et al. (2022), also found a negative association between human and organizational capacity weaknesses and project outcomes and between human and organizational capacity weaknesses and Bank performance.

Skill transfer. Provision of capacity-building and technical support to diverse stakeholders and implementing agencies to strengthen their project management skills and enhance the effectiveness of their activities was linked to moderate improvements in the M&E quality rating, which was 2.88 on average, compared with 2.63 in projects that did not report skill transfer as an implementation factor. Nevertheless, only 15 percent of projects indicated that transfer of skills was predominantly advantageous to the implementation process.

Project-Related Factors

Time allocation or task sequencing. Twenty-seven percent of projects reported challenges related to insufficient time allocation or inappropriate timing and sequence of task to implement project activities, which was linked to slightly lower efficiency (average 2.51 versus 2.67) and Bank performance ratings (average 4.04 versus 4.38).

Beneficiary targeting. Fifteen percent of projects, predominantly in the Human Development Practice Group, reported having appropriate beneficiary targeting, such as targeting lower-income populations, vulnerable groups such as women and marginalized farmers, and schools in rural and remote areas. When beneficiaries are appropriately targeted, project resources are efficiently used, and this in turn can improve supervision and monitoring of project activities because project managers can more easily track progress and make necessary adjustments to ensure that the project achieves its objectives. Additionally, appropriately targeting beneficiaries can build trust and credibility with stakeholders, which can improve the quality of supervision. Projects reporting appropriate beneficiary targeting had slightly higher efficiency (4.83), M&E quality (2.90), and quality of supervision (2.80) ratings than projects that did not (4.54, 2.62, 2.49, respectively).

(continued)

Box D.1. Implementation Factors Statistically Associated with Project Performance Ratings (cont.)

Procurement. About 38 percent of projects reported challenges with procurement management systems, including inefficient contract management and delays, although they only marginally affected project performance ratings. For instance, for projects that reported procurement challenges, the average outcome rating was 4.21, compared with 4.45 for those without such challenges.

Financing mechanism. Nine percent of projects emphasized the importance of selecting the appropriate financing instrument for successful project implementation, such as specific investment loans, trust funds and grants, investment project financing, input-based and disbursement-linked indicators, and results-based financing modality. These projects had, on average, slightly higher outcome, M&E quality, efficiency, and Bank performance ratings than those that did not report financing mechanism as a factor affecting implementation.

Budgeting. Twenty-seven percent of projects reported challenges related to funding gaps associated with inadequate budget provisions that increased project costs, delayed counterpart funding, and decreased co-financing from other sources. These were caused by declines in government revenue, low budgets, lack of timely budget resources, and lengthy government procedures. These challenges of inadequate budgets can limit resources available to project teams, making it difficult for them to oversee project implementation effectively, and delays caused by budget constraints can force project teams to rush to complete work quickly, potentially compromising the quality of supervision. Projects reporting budgeting challenges had lower quality of supervision ratings (average 4.42, in the range of moderately satisfactory) than those that did not (average 4.64, in the range of satisfactory).

Risk identification and mitigation. Forty-seven percent of projects that identified key risks during the project preparation phase, which could affect implementation and outcomes, and outlined mitigation measures to address them had higher performance ratings—average outcome rating of 4.50, compared with 4.24 for projects that did not.

Source: Independent Evaluation Group.

Figure D.12. Distribution of Factors Affecting Implementation According to Outcome Rating

a. Frequency

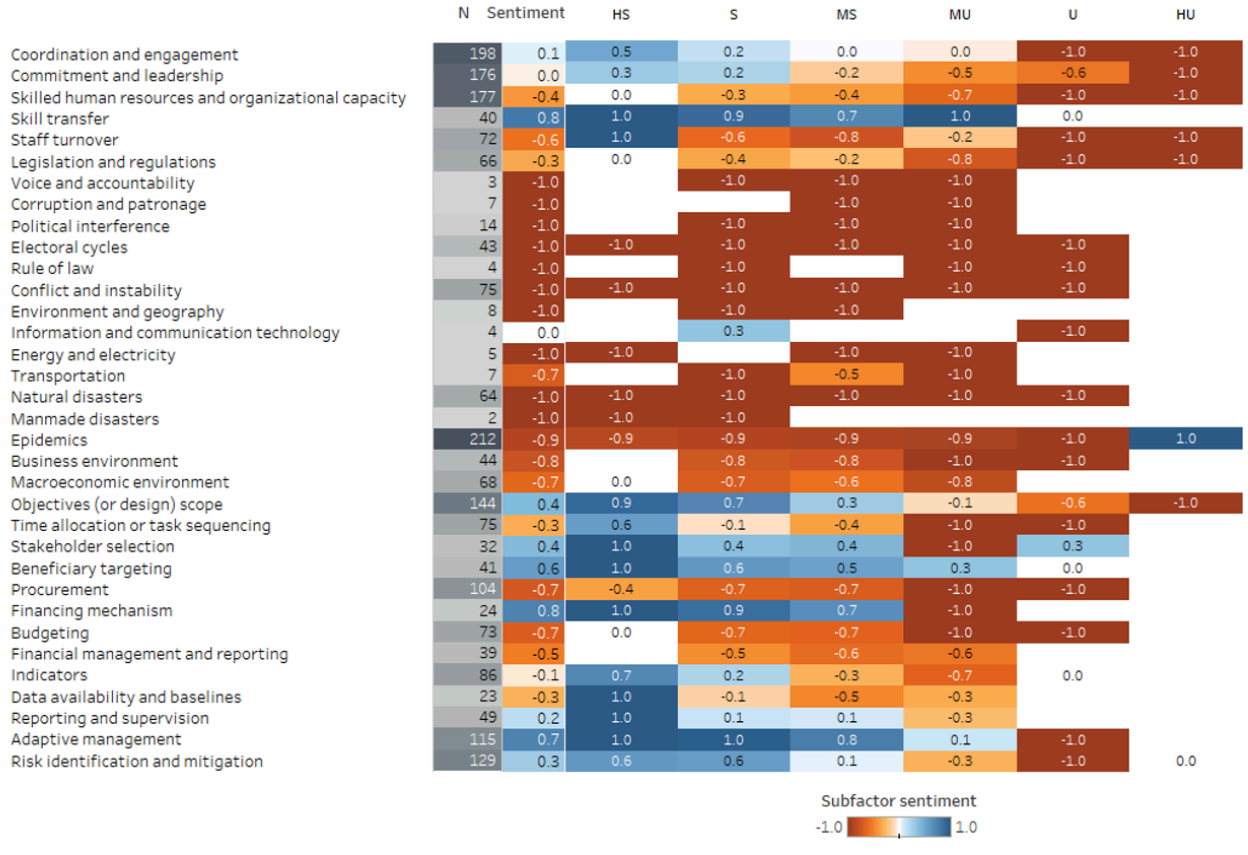
	N	%	20 HS	112 S	97 MS	35 MU	8 U	1 HU
Coordination and engagement	198	73	70	74	73	63	88	100
Commitment and leadership	176	64	65	63	61	71	88	100
Skilled human resources and organizational capacity	177	65	60	59	66	74	100	100
Skill transfer	40	15	10	16	14	11	25	0
Staff turnover	72	26	10	27	27	29	38	100
Legislation and regulations	66	24	30	22	26	23	13	100
Voice and accountability	3	1	0	1	1	3	0	0
Corruption and patronage	7	3	0	0	4	9	0	0
Political interference	14	5	0	3	9	6	0	0
Electoral cycles	43	16	15	13	18	20	13	0
Rule of law	4	1	0	2	0	3	13	0
Conflict and instability	75	27	20	23	31	37	25	0
Environment and geography	8	3	0	4	4	0	0	0
Information and communication technology	4	1	0	3	0	0	13	0
Energy and electricity	5	2	5	0	2	6	0	0
Transportation	7	3	0	2	4	3	0	0
Natural disasters	64	23	15	23	25	26	25	0
Manmade disasters	2	1	5	1	0	0	0	0
Epidemics	212	78	75	74	79	86	75	100
Business environment	44	16	0	15	23	11	13	0
Macroeconomic environment	68	25	10	21	31	34	0	0
Objectives (or design) scope	144	53	55	49	54	57	63	100
Time allocation or task sequencing	75	27	35	22	28	37	38	0
Stakeholder selection	32	12	15	10	14	3	38	0
Beneficiary targeting	41	15	20	18	13	9	13	0
Procurement	104	38	40	29	45	40	75	0
Financing mechanism	24	9	5	14	6	3	0	0
Budgeting	73	27	10	31	25	29	25	0
Financial management and reporting	39	14	0	17	13	20	0	0
Indicators	86	32	35	31	28	43	25	0
Data availability and baselines	23	8	5	6	12	9	0	0
Reporting and supervision	49	18	30	21	11	23	0	0
Adaptive management	115	42	50	44	37	49	38	0
Risk identification and mitigation	129	47	75	54	36	37	63	100

Subfactor tagged

0% 100%

(continued)

b. Sentiment



Source: Independent Evaluation Group.

Note: HS = highly satisfactory; S = satisfactory; MS = moderately satisfactory; MU = moderately unsatisfactory; U = unsatisfactory; HU = highly unsatisfactory.

For projects reporting a particular factor, Pearson and Spearman correlations indicated a positive, statistically significant association between the implementation factor's sentiment and project ratings. In other words, when a factor was identified as facilitating implementation (positive sentiment), project ratings tended to be higher, and when a factor was identified as hindering implementation (negative sentiment), project ratings tended to be lower. This is highlighted in figure D.12, which shows the percentage of projects that reported implementation factors and their overall sentiment across outcome ratings.

All context-related factors, such as legislation and regulation, conflict and instability, natural disasters, epidemics, and macroeconomic environment, were reported as hindering implementation in both successful and unsuccessful projects. However, the impact of stakeholder- and project-related factors on project implementation varied across outcome ratings. Implementation challenges related to coordination and engagement and commitment and leadership were more frequently reported in unsuccessful projects (with moderately unsatisfactory or below outcome ratings). In contrast, successful projects often cited these factors as facilitators of project implementation, indicating a positive sentiment.

Similarly, objectives (or design) scope, data and monitoring issues, adaptive management, and risk identification and mitigation were predominantly reported as facilitators of project implementation in successful projects and as challenges to implementation in unsuccessful projects.

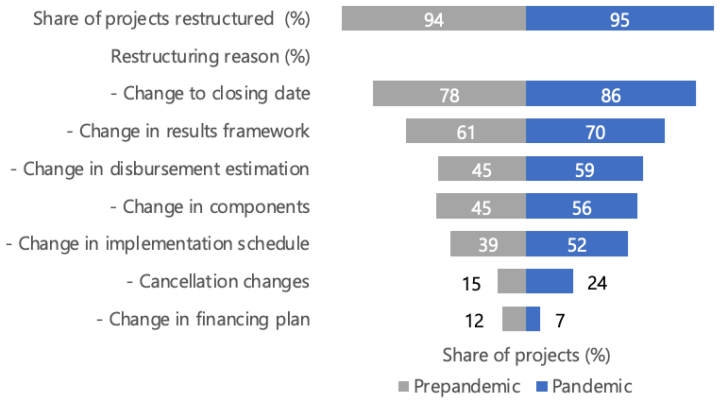
Project Restructuring

To investigate the impact of the COVID-19 pandemic on project restructuring, the analysis focused on various aspects of restructuring, including its occurrence, frequency, and underlying reasons. The analysis encompassed the prepandemic and *RAP 2023* cohorts.

Projects in the *RAP 2023* cohort underwent more restructuring than those in the prepandemic cohort. Although the overall share of projects that underwent restructuring remained relatively stable, with 91 percent of projects experiencing restructuring in the prepandemic cohort and 93 percent

in the *RAP 2023* cohort, there was a notable change in average number of restructurings per project. Before the pandemic, projects had an average of 1.9 restructurings, whereas during the pandemic, this number increased to 2.6 restructurings per project. This increase is statistically significant. The analysis revealed statistically significant shifts in the reasons for project restructuring between the prepandemic and the *RAP 2023* cohorts, specifically regarding implementation scope and process. Notably, the share of restructurings attributed to changes in closing dates, results frameworks, disbursement estimations, components, implementation schedules, and cancellations of financing increased. Conversely, restructurings prompted by changes in financing plans decreased (figure D.13).

Figure D.13. Occurrence and Reasons for Restructuring



Source: Independent Evaluation Group.

In the *RAP 2023* cohort, the analysis found a similar pattern in the reasons for project restructuring. This *RAP* compared the share of restructurings that took place after March 2020 with the project's exposure to COVID-19. For instance, the average ratio between the percentage of restructurings due to fund cancellations after March 2020 and the percentage of project life span that took place after March 2020 reached 3.7, reflecting a rise in restructuring during the pandemic (table D.2). The same pattern was also observed for reasons such as changes in disbursement estimations, implementation schedules, closing dates, components, and results frameworks. Conversely, all restructurings for the *RAP 2023* cohort due to changes in financing plans occurred before March 2020.

The two dominant reasons for restructuring in the *RAP 2023* cohort were changes in project closing dates and results frameworks. According to restructuring data, 86 percent of projects adjusted the closing date, and 70 percent modified the results framework (figure D.13).

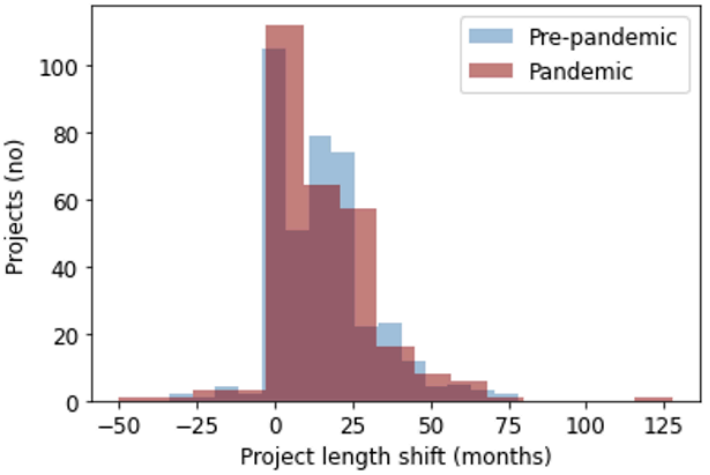
Table D.2. Ratio of Proportion of Restructurings since March 2020 to Share of Project Life Span after March 2020

Restructuring Reason	Ratio
Cancellation changes	3.69
Change to closing date	1.90
Change in disbursement estimation	1.16
Change in implementation schedule	1.12
Change in components	1.04
Change in results framework	1.02
Change in financing plan	0.00

Source: Independent Evaluation Group.

The significant share of projects that underwent restructuring because of changes in closing dates is consistent with the increasing trend of project extensions.[3] The analysis also found an increase in the share of projects that were extended, from 72 percent in the prepandemic cohort to 79 percent in the *RAP 2023* cohort, although there was no significant change in the length of the extension, with mean and median extension periods of 15 months and 12 months in both cohorts (figure D.14.).

Figure D.14. Distribution of Number of Projects According to Change in Project Length (Months)



Source: Independent Evaluation Group.

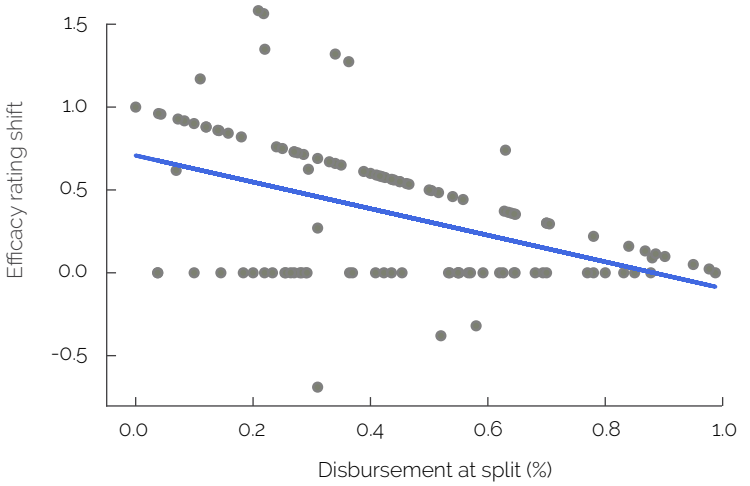
With the increase in projects being restructured because of changes in results frameworks, from 60 percent in the prepandemic cohort to 70 percent in the *RAP 2023* cohort, some of these restructurings led to the adoption of a split rating methodology. Indeed, the percentage of projects with split ratings increased from 3 percent in the prepandemic cohort to 22 percent in the *RAP 2023* cohort.

Independent Evaluation Group staff followed the Implementation Completion and Results Report Review guidelines to independently assess the appropriateness of applying a split rating versus assessing the entire project. A split rating typically applies when (i) project objectives or key associated outcome targets were revised during implementation and (ii) achievements based on original objectives and targets differed from those based on revised objectives and targets. For instance, if the project expanded its scope and the targets for the original geographical area were achieved but those for the newly added geographical areas were not, a split rating was applied. Likewise, when the project scope was decreased through a downward revision of targets and the original target was not achieved but the revised one was achieved, a split rating was applied. A split rating accounts for project achievements measured against both the original and the revised

objectives and targets weighted by the disbursement rate at the time of the revisions when deriving the project's overall efficacy and outcome ratings.

Moreover, the analysis found that projects could benefit from prompt course corrections to adapt to changing circumstances such as the COVID-19 pandemic. A review of all Implementation Completion and Results Report Reviews with split ratings identified a negative correlation between the timing of the split (measured according to the percentage of disbursement) and the shift in efficacy rating (figure D.15). The earlier the split occurred, the more significant the improvement in the rating.

Figure D.15. Timing of Project Revisions and Shift in Efficacy Rating in FY19–22 When a Split Rating Was Applied



Source: Independent Evaluation Group.

The pattern is further illustrated in table D.3, which contrasts the *RAP 2023* cohort's original efficacy ratings with the overall efficacy ratings for two distinct groups: the early revision group (61 percent of projects), for which the split occurred before 50 percent disbursement, and the late revision group (39 percent of projects), for which the split happened after 50 percent disbursement. In the early revision group, a significant majority of projects (38 percent out of the 61 percent) demonstrated an increase in overall efficacy. Conversely, in the late revision group, only a single project (1 percent out of the 39 percent) exhibited improvement.

Table D.3. Timing of Project Revisions and Shift in Efficacy Rating in *RAP 2023* Cohort When a Split Rating Was Applied (percentage of projects)

Efficacy Rating (Original)	Overall Efficacy			
	Negligible	Modest	Substantial	High
Early revised projects		13	50	3
Negligible		10		
Modest		8	48	
Substantial		3	28	5
High				
Late revised projects	3	17	13	
Negligible	10			
Modest		50	5	
Substantial			35	
High				

Source: Independent Evaluation Group.

Note: RAP = Results and Performance of the World Bank Group.

The surge in restructuring spurred by the cancellation of project funds since March 2020 are consistent with findings on the specific ways the pandemic affected project implementation. The content analysis of the epidemic subcategory also found that the pandemic pressured government budgets and forced a realignment of priorities, with projects having to cancel or redirect funds toward pandemic response efforts.

[1] The term “sentiment” refers to the characterization of the language used in the ICR to describe a specific factor as either positively or negatively impacting the implementation of projects (see Appendix A for more details).

[2] Ortega Nieto et al. (2022) used data from the "Delivery Challenges in Operations for Development Effectiveness" (DeCODE) developed by the Global Delivery Initiative. Their study examined project performance and the attainment of development objectives across 42 specific delivery challenges, drawing from a dataset of over 5,000 lending projects spanning the period from 1995 to 2015.

[3] Potential discrepancies between restructuring due to change in closing date and project extension could be attributed to several factors: project extension due to additional financing, which is not reflected in the restructuring data; modification of the loan closing date during the restructuring, which did not affect the project closing date; and quality problems with project operations data.

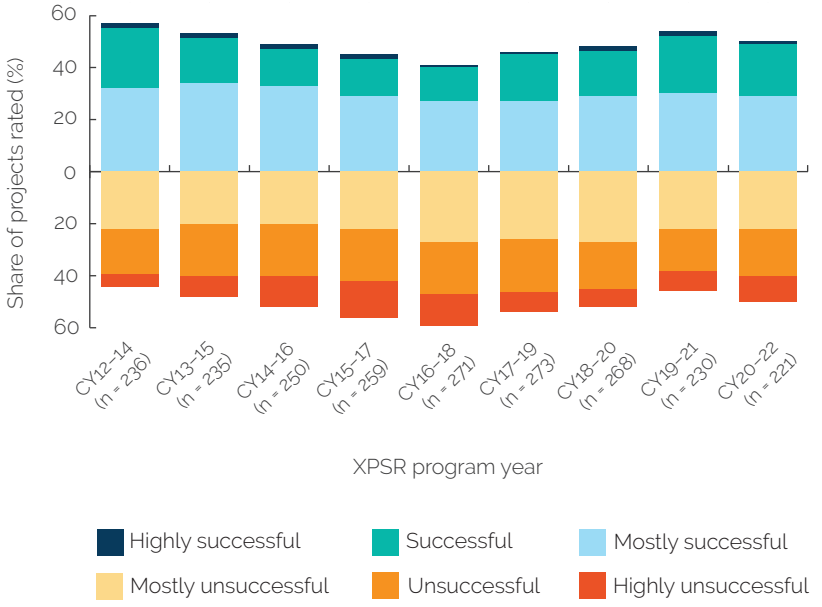
Appendix E. International Finance Corporation Project Rating Trends and Patterns

This appendix presents trends and patterns of International Finance Corporation (IFC) investment and advisory project performance across various project categories, such as industry group, region, country lending group, and country fragile and conflict-affected situation (FCS) status. The reported ratings trends and patterns are as of June 30, 2023. (See appendix A for the methodology.)

Investment Projects

IFC's overall development outcome success ratings for investment projects have been improving since calendar year (CY)16–18 but declined slightly in CY20–22. The percentage of investment projects with outcomes rated mostly successful or better (MS+) decreased from 53 percent in CY19–21 to 50 percent in CY20–22. The decline was driven by the weaker performance of CY22 investment projects, with success rates dropping significantly from 59 percent in CY21 to 50 percent in CY22. The decline in development outcome ratings was mainly due to lower average ratings rather than changes in the evaluated portfolio composition. The decline was also observed in IFC Expanded Project Supervision Report self-ratings. On a granular basis, share of projects with unsuccessful and highly unsuccessful ratings increased, the share with mostly unsuccessful ratings remained the same, and the share at the top end of the development outcome ratings distribution shrank (figure E.1).

Figure E.1. IFC Investment Project Development Outcome Ratings



Source: Independent Evaluation Group XPSR database.

Note: IFC = International Finance Corporation; XPSR = Expanded Project Supervision Report.

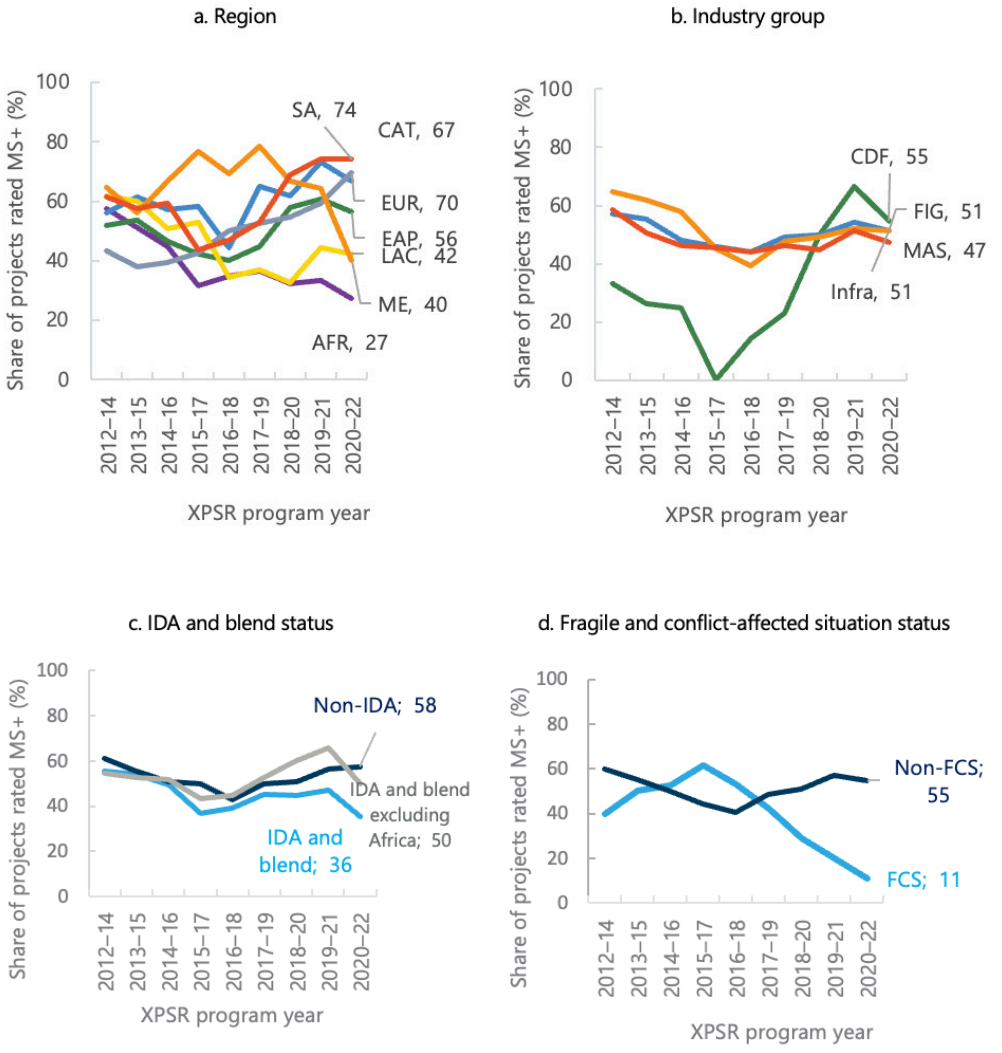
IFC investment projects in the Middle East, Central Asia and Türkiye, and Africa regions experienced the most significant decline in development outcome ratings in CY20–22, driven by adverse macroeconomic factors, effects of the pandemic, and an increase in business risks. The Middle East, and Central Asia and Türkiye projects accounted for a small share of IFC's overall portfolio, whereas African projects accounted for 23 percent of all projects, reducing IFC's average performance ratings. The 2022 *Results and Performance of the World Bank Group* (*RAP*) highlighted the low and declining performance in Africa, which was also observed in this *RAP*. Most investment projects in Africa were in International Development Association (IDA) and blend (eligible for IDA and International Bank for Reconstruction and Development lending) countries, and their performance declined from 33 percent of projects rated MS+ in CY19–21 to 27 percent in CY20–22 (figure E.2, panel a).

The development outcome success ratings of investment projects in Latin America and the Caribbean region, representing 24 percent of the IFC evaluated portfolio, was low, with fewer than half rated MS+. In contrast, performance has been steadily improving in the South Asia, and Europe regions,

with about 70 percent of investment projects rated MS+ for development outcome. Projects in these two regions accounted for 12 percent and 10 percent of the IFC evaluated portfolio, respectively. In East Asia and Pacific, which had the third-largest evaluated portfolio (18 percent of the total), the share of investment projects with outcomes rated MS+ changed only marginally from 61 percent in CY19–21 to 56 percent in CY20–22.

At the industry group level, ratings decline in real sector investment projects in particular contributed to the decrease in overall IFC development outcome ratings (figure E.2, panel b). Success rates for Manufacturing, Agribusiness, and Services investment projects, which accounted for 32 percent of IFC's evaluated portfolio, declined from 51 percent in CY19–21 to 47 percent in CY20–22. For Infrastructure and Natural Resources, which accounts for 19 percent of IFC's evaluated portfolio, the share of investment projects rated MS+ remained about 51 percent in CY20–22. Changes in the portfolio size of Financial Institutions Group and Disruptive Technologies and Funds investment projects offset the change in their average ratings (see the Decomposition Analysis section). The success ratings of Financial Institutions Group investment projects, which accounted for the largest share of IFC's evaluated portfolio (39 percent), declined slightly, to 51 percent in CY20–22 from 54 percent in CY19–21. Accounting for 10 percent of IFC's evaluated portfolio, Disruptive Technologies and Funds investment project success rates declined from 67 percent in CY19–21 to 55 percent in CY20–22.

Figure E.2. IFC Investment Projects Rated Mostly Successful or Better for Select Project Categories



Source: Independent Evaluation Group XPSR database.

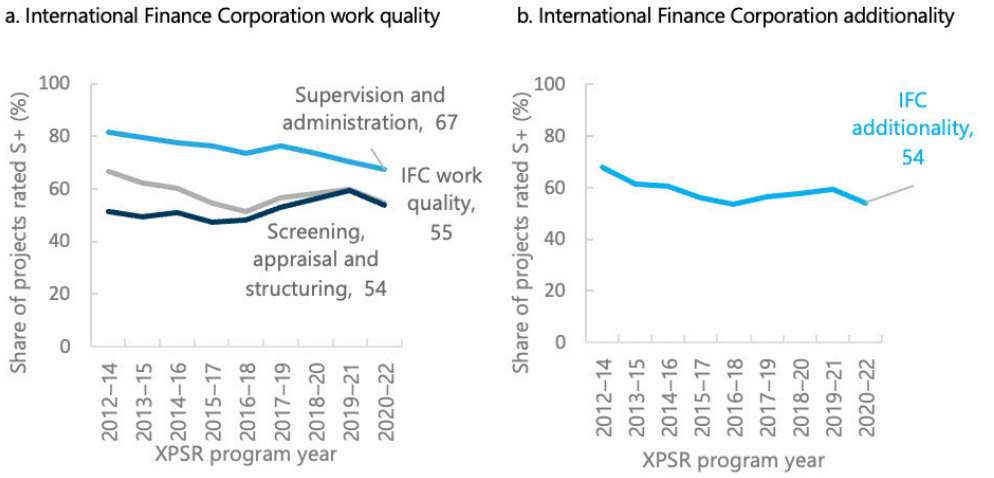
Note: Global and regional projects are not shown in panels c and d. AFR = Africa; CAT = Central Asia and Türkiye; EAP = East Asia and the Pacific; EUR = Europe; LAC= Latin America and the Caribbean; ME = Middle East; SA = South Asia; CDF = Disruptive Technologies and Funds; FIG = Financial Institutions Group; Infra = Infrastructure and Natural Resources; MAS = Manufacturing, Agribusiness and Services; FCS = fragile and conflict-affected situation; IDA = International Development Association; IFC = International Finance Corporation; MS+ = mostly successful or better; XPSR = Expanded Project Supervision Report.

A notable decline in development outcome success ratings for projects in IDA and blend countries contributed to the decline in overall development outcome ratings. The development outcome ratings of investment projects in IDA and blend countries, accounting for 27 percent of IFC's evaluated portfolio, deteriorated, with the share of projects rated MS+ decreasing from 47 percent in CY19–21 to 36 percent in CY20–22 (figure E.2, panel c). Excluding projects in Africa region, the share of IDA and blend investment projects with development outcomes rated MS+ was 50 percent in CY20–22. Projects in non-IDA countries historically outperformed those in IDA and blend countries. Fifty-eight percent of non-IDA investment projects were rated MS+ percent in CY20–22, similar to CY19–21.

Development outcome success ratings of investment projects in FCS countries were low and continued to decline, driven by weak project business performance and low and deteriorating environmental and social performance. The share of FCS investment projects, which accounted for 8 percent of IFC's evaluated portfolio, rated MS+ dropped from 20 percent in CY19–21 to 11 percent in CY20–22 (figure E.2, panel d). This decline was largely due to a decrease in their average ratings rather than changes in their portfolio size. Projects in non-FCS countries outperformed those in FCS countries, but their performance declined, with the share of those rated MS+ decreasing from 57 percent to 55 percent over the same period.

Performance in other dimensions such as IFC work quality, IFC additionality, and investment outcome weakened in CY20–22. In addition to development outcome objectives, it is essential for IFC to deliver high work quality, additionality, and investment outcomes to achieve its corporate purpose. Previous *RAP*s established that development outcome ratings in investment projects are associated with IFC work quality, particularly front-end work quality, that is, screening, appraisal and structuring, and additionality, and investment outcomes are critical for IFC's own financial sustainability. The share of investment projects with high up-front work quality ratings decreased from 59 percent in CY19–21 to 54 percent in CY20–22, whereas supervision work quality ratings stayed around 70 percent over the same period (figure E.3, panel a).

Figure E.3. IFC Investment Project Work Quality and Additionality Ratings



Source: Independent Evaluation Group XPSR database.

Note: IFC = International Finance Corporation; XPSR = Expanded Project Supervision Report.

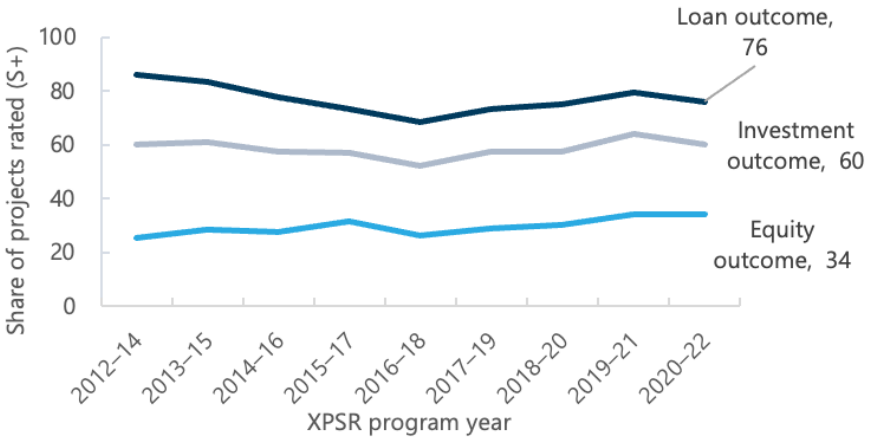
IFC work quality success ratings in African and IDA and blend investment projects were lower than the IFC average and declined in CY20–22, contributing to the overall weakening of IFC work quality. The challenging environment in these countries requires that IFC conduct more thorough due diligence, risk mitigation, and investment structuring on the front end and provide better implementation support during supervision. The share of African investment projects with high work quality ratings decreased from 48 percent in CY19–21 to 35 percent in CY20–22, and IDA and blend project work quality success ratings declined from 57 percent to 47 percent over this period. IFC work quality in investment projects in FCS countries was on par with that in non-FCS countries, with 56 percent rated satisfactory or better in CY20–22.

Overall IFC additionality ratings weakened. The share of investment projects with high additionality ratings was 54 percent in CY20–22, down from 59 percent in CY19–21 (figure E.3, panel b). IFC additionality success ratings in challenging environments were lower than the IFC average. IFC realized its anticipated additionality in 37 percent of African investment projects, 33 percent of FCS projects, and 47 percent of IDA and blend projects. The gap between anticipated and realized additionality in these markets was larger

for nonfinancial additionality than for financial additionality. For example, the gap in provision of knowledge and innovation additionality was 27 percent in Africa, 20 percent in IDA and blend, and 17 percent in FCS countries. The gap in setting new or better standards, for example, in environmental and social, and corporate governance practices, was 16 percent in African, 17 percent in IDA and blend, and 22 percent in FCS countries.

IFC's overall investment outcome success ratings declined, although equity performance has remained stable, albeit at a low level. Financial sustainability is important for individual IFC investment project success and for IFC's sustainability as an investor and institution. IFC's overall investment outcome was satisfactory or better in 60 percent of investment projects in CY20–22, down from 64 percent in CY19–21 owing to the slight decline in loan outcome, some of which was due to prepayments (figure E.4). In contrast, IFC equity outcome success ratings have remained stable, although just about a third of equity investments generated satisfactory returns. A significant share of investment projects in challenging environments, such as African, IDA and blend, and FCS countries, delivered neither high development outcomes nor satisfactory investment returns. IFC's development and investment outcome ratings were both low in 51 percent of African projects, 39 percent of IDA and blend projects, and 56 percent of FCS projects.

Figure E.4. IFC Investment Outcome Ratings



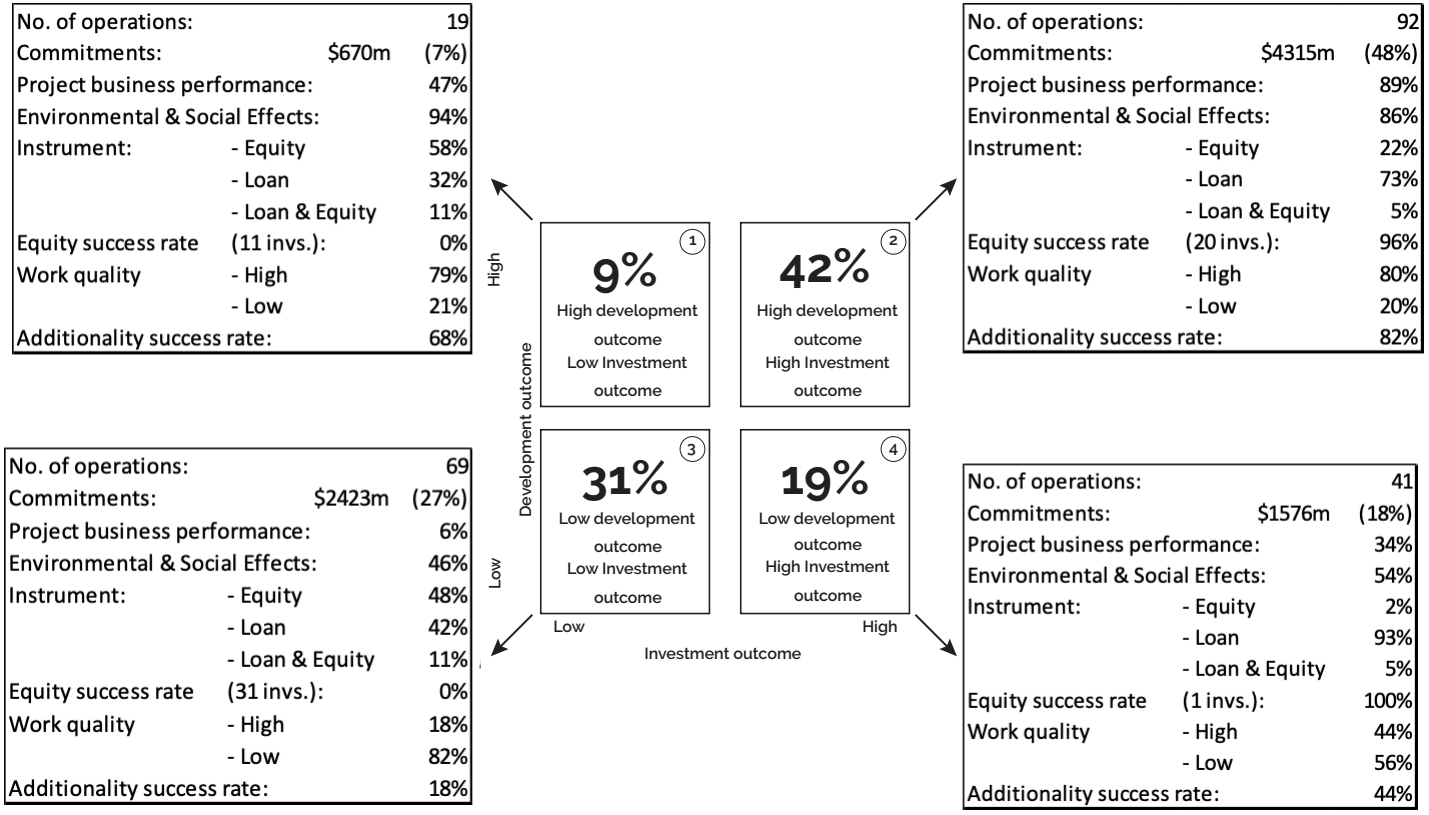
Source: Independent Evaluation Group XPSR database.

Note: IFC = International Finance Corporation; XPSR = Expanded Project Supervision Report.

IFC achieved the double bottom line of high development outcome ratings and high investment returns in 42 percent of projects by delivering high development outcomes while ensuring high investment returns (figure E.5, upper right quadrant). These investment projects had strong project business performance and satisfactory or better environmental and social compliance and were associated with high IFC work quality and additionality. Nine percent of investment projects delivered high development outcomes, despite having low investment returns for IFC (figure E.5, upper left quadrant). More than half of the investment projects in the upper left quadrant of figure E.5 involved IFC equity investments, the low returns of which mirrored weak project business performance, although better performance in other dimensions of development outcome compensated for these shortcomings. These investment projects were also associated with relatively high IFC work quality and additionality.

Nineteen percent of investment projects fell short of delivering expected development outcomes, although they generated satisfactory investment returns to IFC given that all except three involved loan-only investments, with fixed returns to IFC (lower right quadrant). These investment projects generally had weak business performance and environmental and social shortcomings, and IFC work quality and additionality were not in line with operational standards or expectations in more than half of the projects. Close to one-third of all investment projects failed to achieve high development and high investment outcomes (lower left quadrant), and close to half of these projects involved equity investments, all of which failed to generate satisfactory returns. Their business performance was substantially weak, and environmental and social effects were less than satisfactory in more than half of the projects. IFC work quality and additionality were satisfactory in only about one-fifth of projects.

Figure E.5. Characteristics of Projects with Different Development and Investment Outcomes



Source: Independent Evaluation Group.

Decomposition Analysis

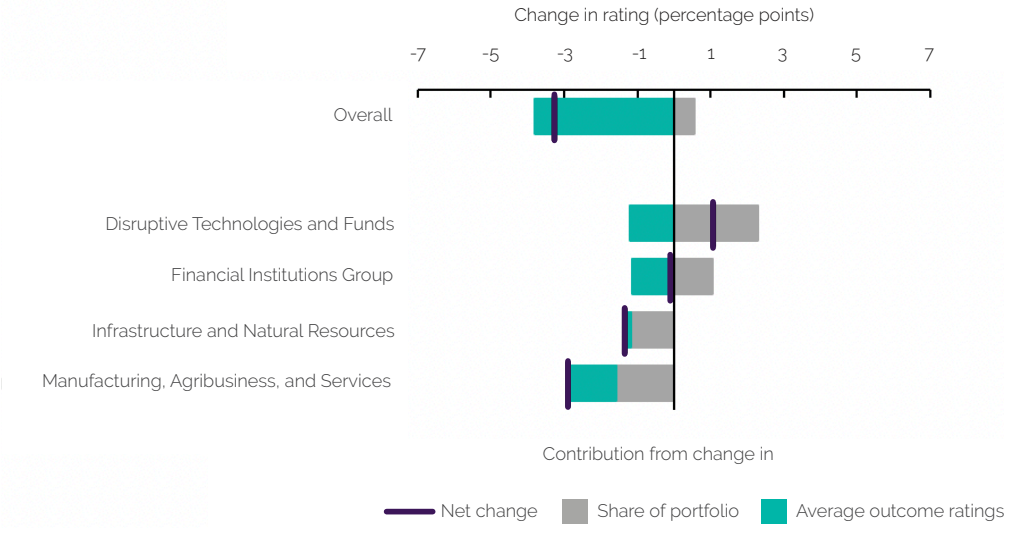
The *RAP* team conducted the decomposition analysis for IFC investment projects, disaggregating the change in average development outcome ratings between the two periods (CY19–21 and CY20–22) into one that is due to changes in ratings of a given project category and another that reflects changes in the share of the projects in that category. The decomposition analysis for IFC projects uses the same methodology as for World Bank projects. The decomposition analysis uses the following formula:

$$\bar{R}_{20-22} - \bar{R}_{19-21} = \sum_i s^i_{20-22} * \left(r^i_{20-22} - r^i_{19-21}\right) + \sum_i r^i_{19-21} * \left(s^i_{20-22} - s^i_{19-21}\right)$$

IFC's overall development outcome success ratings for investment projects declined slightly in CY20–22, with 50 percent of projects rated MS+, compared with 53 percent in CY19–21.

Figure E.6 shows the decomposition of the shift in development outcome ratings according to industry group.

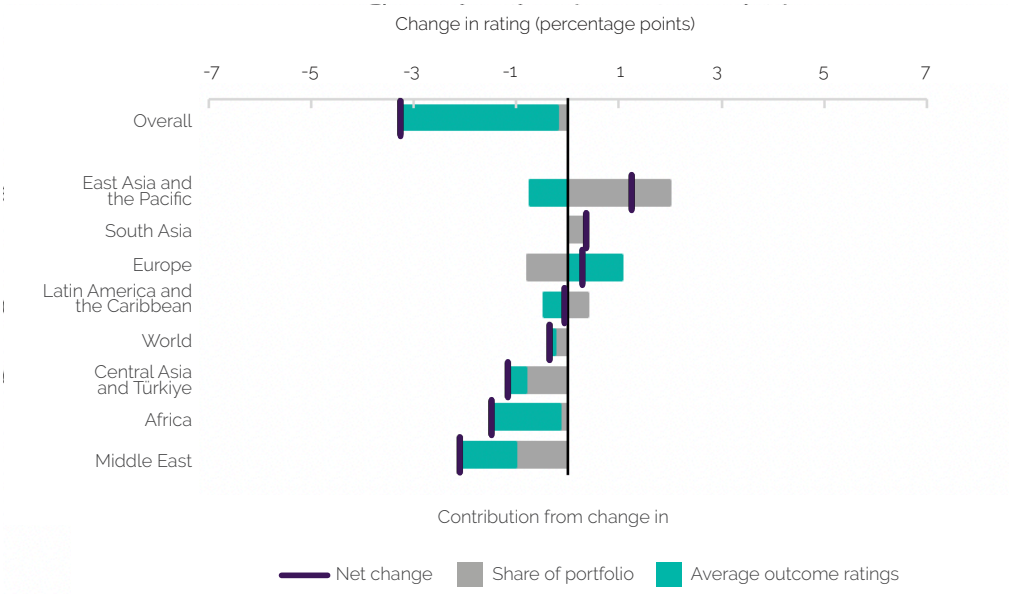
Figure E.6. Decomposition According to Industry Group



Source: Independent Evaluation Group.

Figure E.7 shows the decomposition of the shift in development outcome ratings by region.

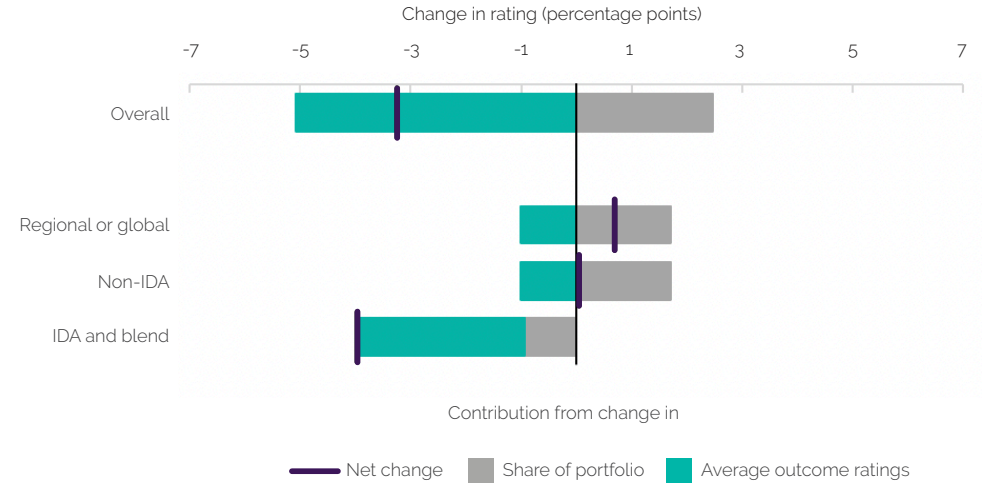
Figure E.7. Decomposition by Region



Source: Independent Evaluation Group.

Figure E.8 shows the decomposition of the shift in development outcome ratings by country lending group.

Figure E.8. Decomposition by Lending Group

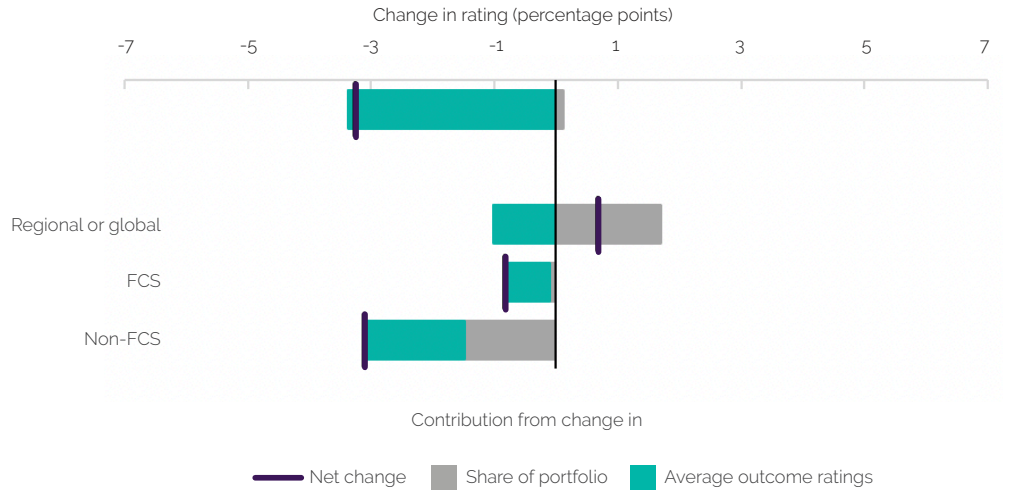


Source: Independent Evaluation Group.

Note: IDA = International Development Association.

Figure E.9 shows the decomposition of the shift in development outcome ratings according to country FCS status.

Figure E.9. Decomposition by Fragile and Conflict-Affected Situation Status



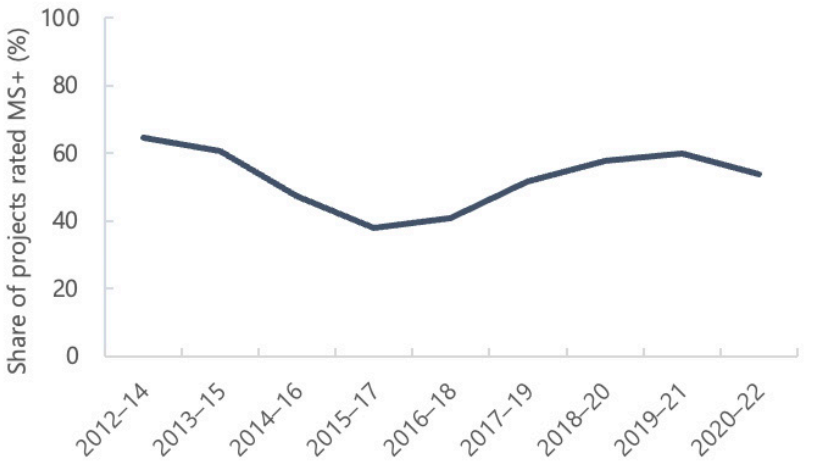
Source: Independent Evaluation Group.

Note: FCS = fragile and conflict-affected situation.

Advisory Projects

The development effectiveness of IFC's advisory projects has been improving since fiscal years (FY)15–17 but declined slightly in FY20–22. The percentage of advisory projects rated MS+ decreased from 60 percent in FY19–21 to 54 percent in FY20–22 (figure E.10), although the decline was not statistically significant. The overall ratings decline was mainly due to a decrease in average ratings rather than changes in the evaluated portfolio composition. (See the decomposition analysis in this appendix.) IFC's self-ratings in Project Completion Reports also declined.

Figure E.10. IFC Advisory Project Development Effectiveness Ratings



Source: Independent Evaluation Group PCR database.

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

Eighty-six percent of advisory projects delivered their direct products and services to clients as expected. However, advisory projects need to show achievement of major intended outcomes and changes in the clients' behaviors and performance. Fifty-four percent of FY20–22 advisory projects achieved ratings of satisfactory or better for outcome achievement by project completion. About one-third of advisory projects had weak strategic relevance and close to half had efficiency shortcomings. However, 23 percent of advisory projects managed to achieve impacts by the project completion.

The declines in several primary business areas pulled down the IFC average development effectiveness ratings in this cohort. Environmental, Social, and Governance; Financial Institutions Group; Public-Private Partnership; and Manufacturing, Agribusiness, and Services projects, which represented large portions of the IFC portfolio, saw declines (see table E.1). The most recent FY22 advisory projects in these primary business areas were challenged by external factors, such as the political conflicts, force majeure events, COVID-19–related disruptions, and client commitment issues, but they also suffered from weaknesses in project preparation and design and shortcomings in monitoring and evaluation. According to advisory services sector highlights in 2020, common challenges for public-private partnership advisory projects that helped governments introduce changes were resistance to

change, limited capacity of counterparts, and underestimation of time and resources needed. Common problems for Financial Institutions Group advisory projects included limited capacity of financial institutions and lack of demand for products or services that the advisory project was launching.

Table E.1. IFC Advisory Project Development Effectiveness Success Ratings by Primary Business Area

Primary Business Area	FY19–21		FY20–22	
	Projects (no.)	Success rate (%)	Projects (no.)	Success rate (%)
Development Economics	2	100	2	100
Disruptive Technologies and Funds	3	33	2	50
Corporate Finance Service	2	0	3	33
Equitable Growth, Finance, and Institutions	49	57	29	62
Economics and Private Sector Development	3	33	2	0
Environment, Social, and Governance	14	71	19	63
Financial Institutions Group	55	71	56	59
Health, Nutrition, and Population	5	20	5	20
Infrastructure and Natural Resources	7	57	9	56
Manufacturing, Agribusiness, and Services	14	50	13	31
Public-Private Partnership	21	67	19	58
Regional Advisory	10	40	16	38
Overall	185	60	175	54

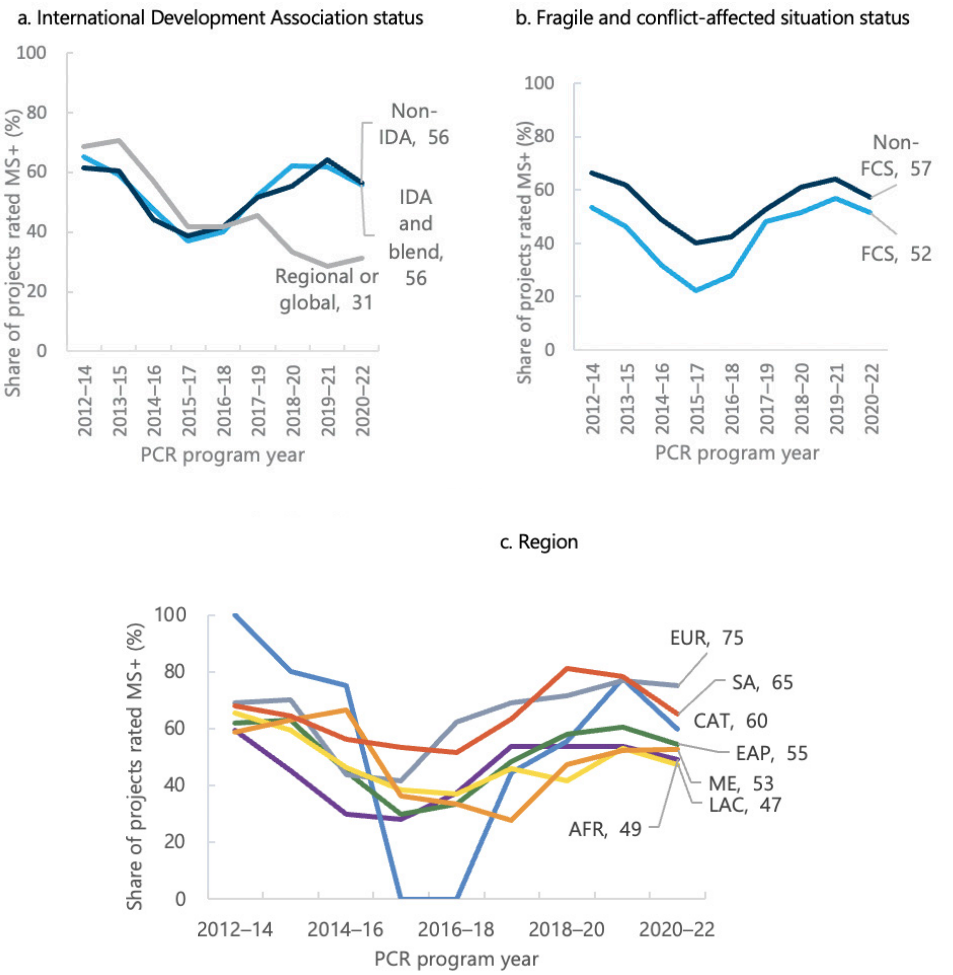
Source: Independent Evaluation Group Project Completion Report database.

Note: IFC = International Finance Corporation.

Development effectiveness success ratings of advisory projects implemented in challenging environments such as IDA and blend, and FCS countries declined in FY20–22. Advisory projects faced challenges from political violence and instability, pandemic disruptions, project design weaknesses,

and shortcomings in monitoring and evaluation. Accounting for nearly 60 percent of the IFC's evaluated portfolio, advisory projects in IDA and blend countries experienced low performance which caused their development effectiveness ratings to decline from 62 percent in FY19–21 to 56 percent in FY20–22 (figure E.11, panel a). The performance of FCS advisory projects, which represented 19 percent of the IFC's evaluated portfolio, declined from 57 percent to 52 percent over the same period (see figure E.11, panel b).

Figure E.11. IFC Advisory Projects Rated Mostly Successful or Better for Select Project Categories



Source: Independent Evaluation Group XPSR database.

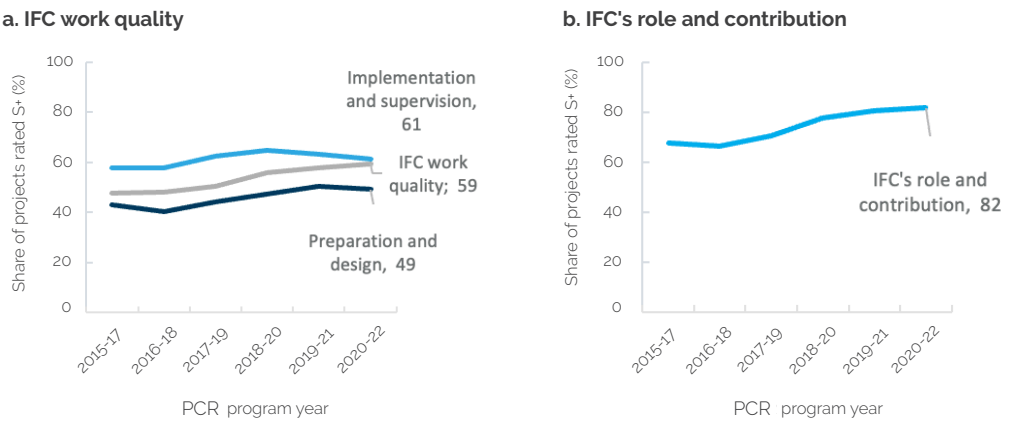
Note: AFR = Africa; CAT = Central Asia and Türkiye; EAP = East Asia and the Pacific; EUR = Europe; LAC = Latin America and the Caribbean; ME = Middle East; SA = South Asia; FCS = fragile and conflict-affected situation; IDA = International Development Association; IFC = International Finance Corporation; MS+ = mostly successful or better; PCR = Project Completion Report.

The development effectiveness rating decline was most significant in the Africa region advisory projects, which made up the largest share of IFC's evaluated portfolio (35 percent of total advisory projects). African advisory projects performed less well than those in other regions, with only about half having high development effectiveness ratings (figure E.11, panel c). Eighty-seven percent of advisory projects in Africa were in IDA and blend countries. Ratings of advisory projects in Central Asia and Türkiye, and South Asia regions also experienced declines. However, the development effectiveness success ratings of projects in these regions were above the IFC average. Advisory projects in the Europe region, which accounted for a small share of the IFC's evaluated portfolio, also performed above the IFC average. In contrast, advisory projects in Latin America and the Caribbean, and the Middle East regions continued to have a low performance, with only about half achieving high development effectiveness. Close to half of advisory projects in the East Asia and the Pacific region also failed to achieve high development effectiveness in FY20–22.

The development effectiveness of advisory projects was highly correlated with IFC work quality, particularly front-end work quality. IFC work quality and role and contribution are vital for realizing high development effectiveness in IFC advisory projects. This has been established in previous *RAP*s and the joint Independent Evaluation Group–IFC Work Quality study in 2017. IFC's preparation and design work quality were satisfactory or better in fewer than half of advisory projects in FY20–22. Although 61 percent of advisory projects had high IFC implementation and supervision work quality, the success ratings declined marginally in FY20–22 (figure E.12, panel a). Project design and preparation ratings were low in advisory projects in the Africa region with the success rate of 43 percent in FY20–22. Likewise, success ratings for IFC's supervision and administration work quality continued to weaken in Africa, and IDA and blend projects. Low development effectiveness ratings of advisory projects were associated with low IFC work quality ratings. Only 21 percent of IDA and blend advisory projects and 15 percent of the Africa advisory projects with low IFC work quality had high development effectiveness. IFC's role and contribution ratings were near historical high levels, with 82 percent of advisory projects rated satisfactory or better in FY20–22. The overall role and contribution success rate remained stable

compared with FY19–21 but was higher than the lowest performance point in FY16–18 (figure E.12, panel b). IFC's role and contribution was generally satisfactory across the entire portfolio of IFC advisory projects evaluated and validated in FY20–22.

Figure E.12. IFC Advisory Project Work Quality, and Role and Contribution Ratings



Source: Independent Evaluation Group PCR database.

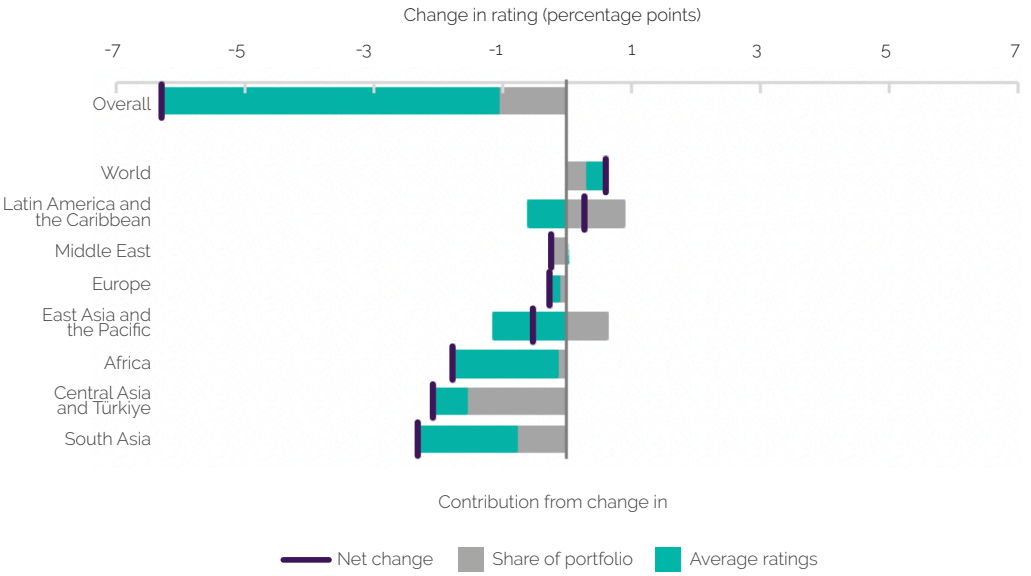
Note: IFC = International Finance Corporation; PCR = Project Completion Report; S+ = satisfactory or better.

Decomposition Analysis

The development effectiveness success ratings of IFC's advisory projects declined slightly in FY20–22, with 54 percent of projects rated MS+, compared with 60 percent in FY19–21.

Figure E.13 shows the decomposition of the shift in development effectiveness ratings by region.

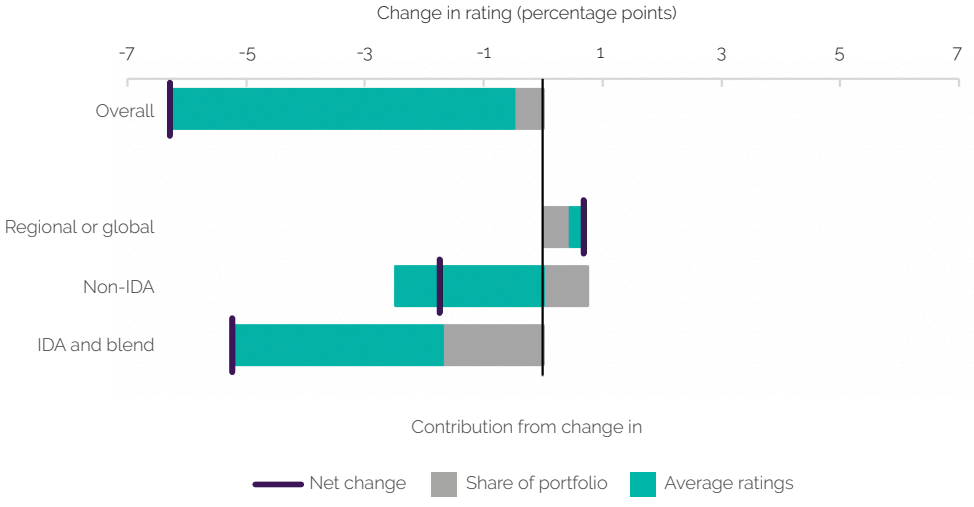
Figure E.13. Decomposition According to Region



Source: Independent Evaluation Group.

Figure E.14 shows the decomposition of the shift in development effectiveness ratings by country lending group.

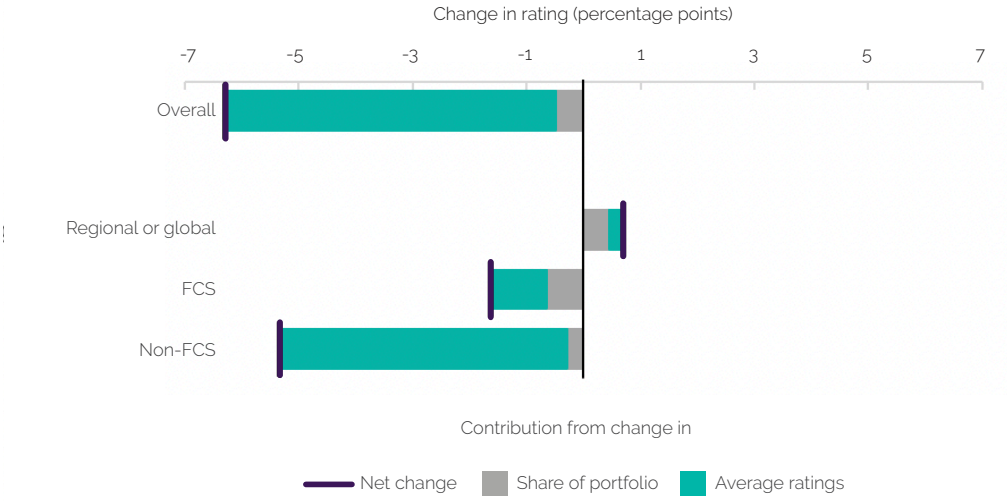
Figure E.14. Decomposition According to Lending Group



Source: Independent Evaluation Group.

Figure E.15 shows the decomposition of the shift in development effectiveness ratings by country FCS status.

Figure E.15. Decomposition According to Fragile and Conflict-Affected Situation Status



Source: Independent Evaluation Group.

Note: FCS = fragile and conflict-affected situation.

Appendix F. Analysis of International Finance Corporation Investment Project Outcome Types

The goal of the outcome type analysis is to describe the intended development outcomes of International Finance Corporation (IFC) investment projects and assess the relationship between those outcomes and performance ratings. This analysis added value because IFC investment projects are not rated at the specific outcome level. This analysis expands on the outcome analysis conducted in the 2021 *Results and Performance of the World Bank Group* (*RAP*) report. For this analysis, the Independent Evaluation Group conducted a deep dive into 170 IFC investment projects evaluated and validated from calendar year (CY)20–22 as of December 31, 2022.

International Finance Corporation Outcome Types

Based on the development outcome statement at approval, an individual project's objective could be classified into several outcome types. IFC's investment project outcomes fall into two broad categories: project-level outcomes and market-level outcomes. Project-level outcomes are defined as a project's direct and indirect effects on stakeholders, the economy, and the environment. Market-level outcomes are derived effects, defined as a project's ability to catalyze systemic changes beyond those effects brought about by the project itself. IFC's 3.0 strategy explicitly prioritizes "creating markets," which falls into the market-level category (IFC 2016, IFC 2019). See box F.1 for examples of project and market outcomes from the *RAP 2023* cohort. Overall, *RAP* identified 33 outcome types (28 project level and 5 foreign investment level) for IFC investment projects that were based on the Anticipated Impact Measurement and Monitoring sector frameworks (box F.1).

For the 170 IFC investment projects covered in the deep-dive analysis, this *RAP* identified 848 individual outcomes. Overall, 100 percent of the investment projects reviewed for the deep-dive analysis pursued project-level

outcomes and 74 percent pursued market-level outcomes. Reviewed projects pursued, on average, 5 different outcomes, consisting of 4 project-level outcomes and 1 market-level outcome.

Box F.1. Examples of IFC Investment Project-Level and Market-Level Outcomes, by Industry Group

Financial Institutions Group

Project-level outcome: Increase in outstanding small and medium enterprise loans, increase in share of microfinance loans, and reduction in nonperforming loans ratio.

Market-level outcome: Demonstration of the viability of lending to microborrowers or small and medium enterprises, deepening of financial markets, and fostering increased competition in the banking sector.

Infrastructure and Natural Resources

Project-level outcome: Increase in renewable energy generation, improvement in information technology infrastructure, increase in access and use of mobile telecommunication services, and number of passengers with access to the road.

Market-level outcome: Diversification of energy mix and increased competition in the information and communication technology sector.

Manufacturing, Agribusiness, and Services

Project-level outcome: Increase in affordable housing supply, increase in purchases from domestic suppliers, increase in quality or affordability of health care services, and increase in tax payments.

Market-level outcome: Demonstration effect on the local agribusiness industry, demonstration of viability of green buildings and promotion of replication, enhanced environment, social, and governance standards to serve as corporate role model.

(continued)

Box F.1. Examples of IFC Investment Project-Level and Market-Level Outcomes, by Industry Group (cont.)

Disruptive Technologies and Funds

Project-level outcome: Percentage of fund investee companies with growth in revenue and returns, increase in job creation at investee companies, and increase in access to information and communication technology services.

Market-level outcome: Demonstration effect through raising of follow-on fund and facilitation of investee companies' emergence as regional players.

Source: Independent Evaluation Group.

Note: IFC = International Finance Corporation.

Prevalence. As shown in table F.1, every IFC investment project pursued the project-level outcome type of improved access to goods and services, which had five subcategories. Other prevalent project-level outcomes were increased employment (55 percent) and quality and affordability of goods and services (48 percent). The most common market-level outcome, increased market competition, was prevalent in 58 percent of all projects. Sustainability in the market, which refers to adoption of climate-related practices, was the second-most prevalent market outcome at 12 percent.

Table F.1. IFC Investment Project Outcome Types: Prevalence

Outcome Type	Outcomes (no.)	Share of Projects (%)
Project-level outcomes		
1.1 - Access to goods and services	242	100
1.1.1 - Access to goods and services (MSME)	50	29
1.1.2 - Access to goods and services (female)	13	8
1.1.3 - Access to goods and services (customers)	88	52

(continued)

Outcome Type	Outcomes (no.)	Share of Projects (%)
1.1.4 - Access to goods and services (miscellaneous)	66	39
1.1.5 - Access to goods and services (direct client level)	25	15
1.2 - Quality and affordability of goods and services	104	48
1.2.1 - Quality of goods and services	37	22
1.2.2 - Affordability of goods and services	53	31
1.2.3 - Improved productivity and efficiency of the direct client	14	8
1.3 - Enhanced capacity of final beneficiaries	23	14
1.4 - Improved living standards (earnings) of individuals	5	3
1.5 - Improved sales and profitability of enterprises	15	9
2.1 - Suppliers and distributors reached	12	7
2.2 - Improved capacity of suppliers and distributors	5	3
2.3 - Improved sales and profitability of suppliers and distributors	28	16
3.1 - Increased employment	94	55
3.2 - Improved capacity and skills	11	6
3.3 - Improved earning of employees	0	0
4.1 - Increased transfers to the government	32	19
5.1 - Increased money spent and transfer to the communities	5	3
6.1 - Enhanced E&S standards of the client	54	32
6.2 - Greenhouse gas reduction	32	19
6.3 - Efficient use of resources	11	6
7.1 - Gross value added	5	3

(continued)

Outcome Type	Outcomes (no.)	Share of Projects (%)
7.2 - Induced or indirect employment	8	5
7.3 - Export sales	4	2
8.1 – Governance	3	2
Total project-level outcomes	693	
Market-level outcomes		
9 - Competition in the market	98	58
10 - Resilience in the market	17	10
11 - Integration in the market	12	7
12 - Inclusiveness in the market	8	5
13 - Sustainability in the market	20	12
Total market-level outcomes	155	
Total outcomes	848	

Source: Independent Evaluation Group.

Note: E&S = environmental and social; IFC = International Finance Corporation.

As shown in table F.2, the prevalence of major outcomes has increased since *RAP 2021*. The shares of projects with intended project-level outcomes related to quality and affordability of goods and services, GHG reduction, and increased employment, as well as market-level outcomes related to enhancing sustainability were considerably higher in CY20–22 than in CY12–19.

Table F.2. Comparison of Prevalence of IFC Investment Project Outcomes

	CY12-19 (*RAP 2021*) n=236		CY20-22 (*RAP 2023*) n=170	
Outcome Level	Outcomes (no.)	Share of Projects (%)	Outcomes (no.)	Share of Projects (%)
Project-level (top 10)				
Access to goods and services**	193	68	242	100
-Access to goods and services (customers)**	97	40	88	52
-Access to goods and services (miscellaneous)	n.a.	n.a.	66	39
-Access to goods and services (MSME)	97	32	50	29
-Access to goods and services (direct client level)	n.a.	n.a.	25	15
Increased employment**	56	24	94	55
Quality and affordability of goods and services**	16	7	104	48
Increased transfers to the government**	9	4	32	19
Greenhouse gas reduction*	30	13	32	19
Improved sales and profitability of suppliers and distributors	29	12	28	16
Enhanced capacity of final beneficiaries	31	13	23	14

(continued)

Outcome Level	CY12-19 (*RAP 2021*) n=236		CY20-22 (*RAP 2023*) n=170	
	Outcomes (no.)	Share of Projects (%)	Outcomes (no.)	Share of Projects (%)
Market-level				
Competition in the market	126	53	98	58
Sustainability in the market**	13	6	20	12
Resilience in the market	24	10	17	10
Integration in the market	28	12	12	7
Inclusiveness in the market	16	6	8	5

Source: Independent Evaluation Group.

Note: IFC = International Finance Corporation; MSME = micro, small, and medium enterprise. n.a. = not applicable.
*Statistically significant at 90% C.I. **Differences statistically significant at 95% C.I.

IFC's four industry groups are as follows: Disruptive Technologies and Funds; Financial Institutions Group; Infrastructure and Natural Resources; and Manufacturing, Agribusiness, and Services. These industry groups tended to support certain outcome types. As shown in table F.3, all four industry groups frequently implemented investment projects intended to increase access to goods and services. Within this broad category,

- Disruptive Technologies and Funds most frequently implemented investment projects to increase access to goods and services at the direct client level;
- Financial Institutions Group to increase access to goods and services for MSMEs;
- Infrastructure and Natural Resources to increase access to goods and services for a variety of client types; and
- Manufacturing, Agribusiness, and Services to increase access to goods and services for customers.

Almost all industries also frequently implemented investment projects intended to increase employment and the quality or affordability of goods and services. All industry groups supported market-level outcomes, with increased market competition being the most prevalent one. Disruptive Technologies and Funds and Financial Institutions Group investment projects also focused on sustainability in the market, while Infrastructure and Natural Resources investment projects focused on resilience in the market and integration in the market. Finally, Manufacturing, Agribusiness, and Services investment projects also pursued the market-level outcome of integration in the market.

Table F.3. Outcome Type by Sector as Share of Projects: Prevalence

Outcome Type	CY20–22 (*n* = 170)			
	CDF (*n* = 14)	FIG (*n* = 71)	INR (*n* = 32)	MAS (n = 53)
Project-level outcomes				
1.1 - Access to goods and services	100	100	100	100
1.1.1 - Access to goods and services (MSME)	36	59	0	6
1.1.2 - Access to goods and services (female)	0	18	0	0
1.1.3 - Access to goods and services (customers)	14	49	34	75
1.1.4 - Access to goods and services (miscellaneous)	57	34	66	25
1.1.5 - Access to goods and services (direct client level)	86	3	22	8
1.2 - Quality and affordability of goods and services	86	38	53	47
1.2.1 - Quality of goods and services	7	1	34	45
1.2.2 - Affordability of goods and services	43	28	25	36
1.2.3 - Improved productivity and efficiency of the direct client	0	1	16	15

(continued)

Outcome Type	CY20–22 (*n* = 170)			
	CDF (*n* = 14)	FIG (*n* = 71)	INR (*n* = 32)	MAS (n = 53)
1.3 - Enhanced capacity of final beneficiaries	36	14	3	13
1.4 - Improved living standards (earnings) of individuals	0	1	3	6
1.5 - Improved sales and profitability of enterprises	64	1	3	8
2.1 - Suppliers and distributors reached	7	0	3	19
2.2 - Improved capacity of suppliers and distributors	0	0	3	8
2.3 - Improved sales and profitability of suppliers and distributors	0	0	13	45
3.1 - Increased employment	71	27	69	81
3.2 - Improved capacity and skills	0	0	9	15
3.3 - Improved earning of employees	0	0	0	0
4.1 - Increased transfers to the government	0	0	41	36
5.1 - Increased money spent and transfer to the communities	0	0	16	0

(continued)

Outcome Type	CY20–22 (*n* = 170)			
	CDF (*n* = 14)	FIG (*n* = 71)	INR (*n* = 32)	MAS (n = 53)
6.1 - Enhanced E&S standards of the client	50	30	28	32
6.2 - Greenhouse gas reduction	29	15	34	11
6.3 - Efficient use of resources	0	1	0	19
7.1 - Gross value added	0	0	3	8
7.2 - Induced or indirect employment	7	0	13	6
7.3 - Export sales	0	3	0	**4**
8.1 - Governance	14	0	3	0
Market-level outcomes				
9 - Competition in the market	93	45	66	60
10 - Resilience in the market	0	10	16	9
11 - Integration in the market	0	0	16	13
12 - Inclusiveness in the market	0	4	3	8
13 - Sustainability in the market	21	17	3	8

Source: Independent Evaluation Group.

Note: E&S = environmental and social; CDF = Disruptive Technology and Funds; FIG = Financial Institutions Group; INR = Infrastructure and Natural Resources; MAS = Manufacturing, Agribusiness, and Services; MSME = micro, small, and medium enterprise.

Achievement. Outcome achievement rates were substantially lower than what was envisaged at approval. IFC investment projects had 848 outcomes, of which 382 were fully achieved (45 percent) and 189 were partially achieved (22 percent). IFC investment projects had 693 project-level outcomes, of which 45 percent were fully achieved and 23 percent were partially achieved. However, there was wide variance in achievement levels among specific outcomes (see table F.4).

Among the more prevalent project-level outcomes (with 15 or more incidences), improved sales and profitability of enterprises had the highest achievement rate at 67 percent. Other prevalent project-level outcomes with high achievement rates included enhanced capacity of final beneficiaries at 57 percent, greenhouse gas reduction at 63 percent, and enhanced E&S [environmental and social] standards of the client at 52 percent. Major project-level outcomes with relatively low achievement rates included increased transfers to the government at 31 percent, improved sales and profitability of suppliers and distributors at 39 percent, and increased employment at 40 percent.

IFC investment projects also pursued 155 market-level outcomes, of which 43 percent were fully achieved and 21 percent were partially achieved. Considering that the market-level achievement rate was almost as high as that of project-level outcomes, the *RAP*'s hypothesis that market-level outcomes were more difficult to achieve than project-level outcomes was not fully confirmed. Resilience in the market was the market-level outcome with the highest achievement rate, at 47 percent, while competition in the market (the most prevalent claim) had an achievement rate of only 45 percent.

Table F.4. IFC Outcome Type Performance: Achievement Rate

Outcome Type	Outcomes (no.)	Outcome Achieved (fully; %)	Outcome Achieved (partially; %)	Outcome Achieved (fully + partially; %)
Project-level outcomes				
1.1 - Access to goods and services	242	44	30	74
1.1.1 - Access to goods and services (MSME)	50	46	22	68
1.1.2 - Access to goods and services (female)	13	62	15	77
1.1.3 - Access to goods and services (customers)	88	43	32	75
1.1.4 - Access to goods and services (miscellaneous)	66	39	33	72
1.1.5 - Access to goods and services (direct client level)	25	44	36	80
1.2 - Quality and affordability of goods and services	104	47	14	61
1.2.1 - Quality of goods and services	37	46	14	60
1.2.2 - Affordability of goods and services	53	47	13	60
1.2.3 - Improved productivity and efficiency of the direct client	14	50	21	71

(continued)

Outcome Type	Outcomes (no.)	Outcome Achieved (fully; %)	Outcome Achieved (partially; %)	Outcome Achieved (fully + partially; %)
1.3 - Enhanced capacity of final beneficiaries	23	57	13	70
1.4 - Improved living standards (earnings) of individuals	5	0	0	0
1.5 - Improved sales and profitability of enterprises	15	67	27	94
2.1 - Suppliers and distributors reached	12	50	25	75
2.2 - Improved capacity of suppliers and distributors	5	40	20	60
2.3 - Improved sales and profitability of suppliers and distributors	28	39	32	71
3.1 - Increased employment	94	40	23	63
3.2 - Improved capacity and skills	11	45	27	72
3.3 - Improved earning of employees	0	n.a.	n.a.	n.a.
4.1 - Increased transfers to the government	32	31	28	59
5.1 - Increased money spent and transfer to the communities	5	60	0	60
6.1 - Enhanced E&S standards of the client	54	52	17	69

(continued)

Outcome Type	Outcomes (no.)	Outcome Achieved (fully; %)	Outcome Achieved (partially; %)	Outcome Achieved (fully + partially; %)
6.2 - Greenhouse gas reduction	32	63	6	69
6.3 - Efficient use of resources	11	55	9	64
7.1 - Gross value added	5	40	60	100
7.2 - Induced and indirect employment	8	25	0	25
7.3 - Export sales	4	50	0	50
8.1 - Governance	3	67	0	67
Total project-level outcomes	693	45	23	68
Market-level outcomes				
9 - Competition in the market	98	45	20	65
10 - Resilience in the market	17	47	29	76
11 - Integration in the market	12	33	25	58
12 - Inclusiveness in the market	8	38	13	50
13 - Sustainability in the market	20	40	20	60
Total market-level outcomes	155	43	21	65
Total outcomes	848	45	22	67

Source: Independent Evaluation Group.

Note: Of project-level outcomes, 8 percent were considered not achieved because the results could not be verified. Of market-level outcomes, 7 percent were considered not achieved because their results could not be verified. E&S = environmental and social; IFC = International Finance Corporation; MSME = micro, small, and medium enterprise.

In comparison with the *RAP 2021* cohort, the achievement rates for most project-level and market-level outcome types were lower in the *RAP 2023* cohort (table F.5). The difference in achievement rates is partly explained by different outcome measurement techniques between the *RAP 2021* and *RAP 2023* teams. *RAP 2021* identified outcomes by referring to back-filled Anticipated Impact Measurement and Monitoring claims and based its assessments on the assessment of Evaluative Notes as well as the *RAP* team's judgment, while *RAP 2023* identified each project-level and market-level outcome by reading through project Evaluative Notes and used the Evaluative Notes author's judgment on the extent of their achievement. Moreover, the sample size for *RAP 2023* (n = 170) was smaller than that for *RAP 2021* (n = 236), which may have created statistical differences. This is borne out by the fact that only the achievement rates for increased employment and improved sales and profitability of suppliers and distributors were statistically significant between the two *RAP* cohorts.

Table F.5. Comparison of Outcome Type Performance: Achievement Rate

Outcome Level	CY12–19 (*RAP 2021*; *n* = 236)		CY20–22 (*RAP 2023*; *n* = 170)	
	Total outcomes (no.)	Outcomes achieved (%)	Total outcomes (no.)	Outcomes achieved (%)
Project-level outcomes (top 10)				
1.1 - Access to goods and services	193	51	242	44
1.1.3 - Access to goods and services (customers)	97	53	86	43
1.1.4 - Access to goods and services (miscellaneous)	n.a.	n.a.	66	39
1.1.1 - Access to goods and services (MSME)	97	51	50	46
1.1.5 - Access to goods and services (direct client level)	n.a.	n.a.	24	44
3.1 - Increased employment**	56	57	94	40
1.2 - Quality and affordability of goods and services	16	63	104	47
4.1 - Increased transfers to the government	9	44	32	31
6.2 - Greenhouse gas reduction	30	70	32	63
2.3 - Improved sales and profitability of suppliers and distributors**	29	66	28	39
1.3 - Enhanced capacity of final beneficiaries	31	45	23	57

(continued)

Outcome Level	CY12–19 (*RAP 2021*; *n* = 236)		CY20–22 (*RAP 2023*; *n* = 170)	
	Total outcomes (no.)	Outcomes achieved (%)	Total outcomes (no.)	Outcomes achieved (%)
Market-level outcomes				
9 - Competition in the market	126	36	98	45
13 - Sustainability in the market	13	38	20	40
10 - Resilience in the market	24	63	17	47
11 - Integration in the market	28	43	12	33
12 - Inclusiveness in the market	16	69	8	38

Source: Independent Evaluation Group.

Note: MSME = micro, small, and medium enterprise.
**Differences statistically significant at 95% C.I.

Inadequate results measurement indicators and lack of evidence prevented verification of achievement of some intended outcomes. Monitoring data were not available for a significant number of outcomes, confirming the *RAP*'s hypothesis that IFC's results measurement indicators are not fully adequate to measure the achievement of outcomes. Given that most of these investment projects were not subject to an Anticipated Impact Measurement and Monitoring assessment at their approval, they continued to be monitored in the Development Outcome Tracking System (DOTS). However, of the 848 total outcomes, 490 did not have monitoring data. This meant that 58 percent of all outcomes were missing monitoring data. Of the 490 outcomes not tracked, 339 were project-level outcomes and 151 were market-level outcomes. In many cases, either IFC or the Independent Evaluation Group used other information sources to validate the outcome.

Some outcomes could not be verified because of lack of appropriate results measurement indicators and evidence, which depressed outcome achievement rates. Of the total outcomes, 8 percent could not be verified because of a lack of evidence and were coded as "cannot be verified", including 8 percent of project-level outcomes and 7 percent of market-level outcomes. Of the 65 outcomes that could not be verified, 91 percent were not tracked by IFC in any monitoring system. Some of the most common reasons an outcome could not be verified were as follows: (i) the project did not have an indicator to track the outcome, (ii) the client did not report relevant information, (iii) there was insufficient evidence to measure achievement, (iv) there was no clarity in how to measure the outcome, (v) the result could not be attributed to the project, or (vi) it was too early to tell.

IFC Outcome Ratings Analysis

Investment project development outcome ratings clearly reflected the actual achievement of specific project outcomes, while also being conditioned by project business, environmental, and social performance. Based on Expanded Project Supervision Report guidelines, the development outcome ratings of IFC investment projects are assigned at the project level and subdimension level but not at the project outcome level. Therefore, this *RAP* expanded the *RAP 2021* outcome analysis by linking the achievement of project outcomes with individual project development outcome ratings. This was an attempt to

answer evaluation question 2, "What has been the evolution of development outcomes pursued, measured, and achieved at the project level, and what is the relationship between outcomes and project performance ratings?"

The *RAP*'s outcome analysis showed that IFC investment projects that achieved more of their outcomes had higher development outcome ratings. Moreover, investment projects that achieved their market-level objectives also achieved higher development outcome ratings (table F.6). For the entire cohort of 170 investment projects, development outcome ratings declined in tandem with lower outcome achievement (at both project level and market level). Project Business Performance (PBP) and environmental and social effects are components of development outcome; therefore, lower PBP and environmental and social performance were also associated with lower investment project development outcome ratings.

Table F.6. IFC Investment Project Development Outcome Ratings and Underlying Outcome Achievement Rates

Development Outcome Rating	Total Projects (no.)	Total Outcomes (no.)	Overall Weighted Achievement Rate (%)	Project-Level Outcome Weighted Achievement Rate (%)	Market-Level Outcome Weighted Achievement Rate (%)	Project Business Performance Average Rating	Environmental and Social Effects Average Rating
Highly successful	3	18	100	100	100	4.0	3.0
Successful	35	161	87	86	90	3.4	3.0
Mostly successful	46	250	73	75	68	2.8	2.8
Mostly unsuccessful	39	198	49	50	45	2.2	2.5
Unsuccessful	32	139	23	25	9	1.4	2.5
Highly unsuccessful	15	82	8	8	8	1.1	1.8

Source: Independent Evaluation Group.

Note: Outcome achievements in projects are measured with the following weights: outcome achieved = 1, partially achieved = 0.5, not achieved = 0, cannot verify = 0. Project business performance and environment and social effects ratings' numerical values are as follows: excellent = 4, satisfactory = 3, partly unsatisfactory = 2, unsatisfactory = 1. IFC = International Finance Corporation.

Highly successful investment projects achieved 100 percent of their total outcomes, including 100 percent of their market-level outcomes. Development outcome ratings declined in tandem with lower outcome achievement. Successful investment projects achieved 87 percent of their total outcomes and 90 percent of their market-level outcomes. Conversely, unsuccessful investment projects achieved only 23 percent of their total outcomes, and 9 percent of their market-level outcomes, and highly unsuccessful investment projects achieved only 8 percent of their total outcomes and 8 percent of their market-level outcomes. There was a clear link between outcome achievement, and especially market-level outcome achievement, and development outcome rating.

An investment project's level of outcome achievement was the main difference in influencing the development outcome rating for borderline projects. Outcome achievement was the main difference between borderline projects (mostly successful and mostly unsuccessful): 73 percent versus 49 percent. Mostly successful projects achieved 68 percent of market-level outcomes, compared with only 45 percent for mostly unsuccessful projects. Financial performance was also an important factor for borderline projects, as the difference in their PBP rating was one full rating difference (satisfactory versus partly unsatisfactory). Mostly successful projects had an average PBP rating of 2.8 (closer to satisfactory than partly unsatisfactory), and mostly unsuccessful projects had an average PBP rating of 2.2 (closer to partly unsatisfactory than satisfactory).

Appendix G. Analysis of Factors Affecting International Finance Corporation Investment Project Implementation and Performance

The goal of this appendix is to describe the factors of success or failure that affected International Finance Corporation (IFC) investment project implementation and performance in the COVID-19 pandemic context and uncover patterns across projects and country characteristics based on a qualitative review.

Exposure to COVID-19 Pandemic

The COVID-19 pandemic presented various unforeseen implementation challenges for IFC investment projects. All investment projects in the combined calendar year (CY)20–22 sample cohort were evaluated and validated during the pandemic after March 2020. Most investment projects in the CY20–22 cohort were affected by COVID-19 to some extent, given that 75 percent of them were still active in IFC portfolio at the time of evaluation. On average, the 170 IFC investment projects covered by the deep-dive factor analysis were exposed to COVID-19 for 24 percent of their active project lives.

Factors Affecting the IFC Investment Project Implementation and Performance in the COVID-19 Context

Leveraging the existing taxonomy of performance factors consisting of 5 categories and 51 subcategories, *RAP 2023* conducted a deep-dive analysis of 170 IFC investment projects validated from CY20–22 in which it identified the top three critical performance factors for each project. Each factor can have either a negative or positive influence on the project performance.

The six most prevalent factors supporting the performance of CY20–22 investment projects were (i) technical expertise and track record of sponsors and clients (30 percent of projects), (ii) favorable business aspects (9 percent), (iii) high market share of the clients' business (7 percent), (iv) favorable technology choices (6 percent), (v) collaboration and coordination among IFC investment and advisory teams (5 percent) and (vi) strong capacity, capitalization, and leverage of sponsors (5 percent). All these factors could be considered within IFC and sponsors' control, and they featured more prominently in high-performing projects. Among these top positive factors, the ability, technical expertise, and track record was a common supporting factor for investment projects across all industries, while the significance of other factors tended to differ by industry. For example, market share, collaboration and coordination within IFC investment and advisory services teams, and capacity, capitalization, and leverage factors were more common for Financial Institutions Group investment projects, while business aspects as a positive factor featured more in Infrastructure, and Natural Resources (INR) and Manufacturing, Agribusiness, and Services (MAS) investment projects, and technology was typically found in INR projects.

The top 5 factors that negatively affected the overall investment portfolio performance in CY20–22 were (i) unfavorable economic issues (23 percent of all investment projects), (ii) high business risks (17 percent), (iii) epidemics and COVID-19 (14 percent), (iv) higher than expected competition (14 percent), and (v) limited technical expertise and track record of sponsors and clients (13 percent). Because economic recessions and the COVID-19 pandemic were not foreseeable, they could be said to have been out of the control of IFC and the sponsor. These negative factors affected the performance of projects across all industries, though the share of affected projects varied for each industry. Economic issues and epidemics and COVID-19 factors affected a relatively large share of Financial Institutions Group, MAS, and Disruptive Technologies and Funds investment projects. Competition affected more INR and MAS real sector investment projects. Technical expertise and track record was more of an issue for Disruptive Technologies and Funds, INR, and MAS investment projects, while business factors affected all industry projects to a similar extent. Box G.1 provides examples of how the factors affected investment projects in various industry groups.

Box G.1. Examples of Supporting and Constraining Factors Affecting IFC Investment Project Performance, by Industry Group

Financial Institutions Group Projects

Supporting factors. Technical expertise and track record as a positive factor meant that the management of financial institutions was experienced and the financial institution had a historical strong performance in terms of earnings, asset quality, and risk management. Market share typically meant that the financial institution was the leader in the respective market (for example, banking, small and medium enterprise lending, microfinance, and housing), which gave it an edge over the competition. "Collaboration and coordination within the International Finance Corporation: advisory services and investment services" typically referred to joint International Finance Corporation investment services and advisory projects that helped improve the capacity of financial institutions, especially in the area of micro, small, and medium enterprise lending. Commitment and motivation meant that the financial institution was not sufficiently committed to lending to the target beneficiaries, that is, project objectives were not aligned with the financial institution strategy.

Constraining factors. Business factors meant that the financial institutions moved away from lending to the target beneficiaries (for example, micro, small, and medium enterprises; agribusiness; and affordable housing) as a result of the higher-risk profile. In a more adverse environment, the financial institutions turned toward less riskier loans to corporates or investing in government securities. This risk management was needed to help preserve capital, however, the consequence was that the development impact of these projects was reduced. Asset quality could mean that the financial institution did not sufficiently provision for bad loans or had deficiencies in credit risk management. Legal or regulatory factors meant that the regulatory environment became more adverse during project implementation, with interest rate caps or new policy requirements on financial institutions.

(continued)

Box G.1. Examples of Supporting and Constraining Factors Affecting IFC Investment Project Performance, by Industry Group (cont.)

Infrastructure and Natural Resources Projects

Supporting factors. Legal or regulatory factors meant that the projects benefited from effective structuring of concession agreement and supportive government policies and initiatives. Technology meant that the projects benefited from technically and commercially viable technology, with an edge over inefficient or costlier options. Pricing meant that the projects benefited from favorable tariffs or upward trends in market prices of their products.

Constraining factors. Business factors meant that the projects had flawed, untested, or fragile business models or experienced slowdown in market growth. Legal or regulatory factors meant that the projects were affected by failure to obtain the required licenses, an unexpected government decision to withhold value-added tax reimbursements, and disputes between the government and project company regarding the curtailment of fuel supply. Political factors meant the issues related to illiquid public sector offtaker with payment dependence on the government, inability of the government to meet its obligations in terms of fuel supply and offtake payments, delay in the commissioning attributed to the government, and regulatory changes because of the government suspension of the privatization program.

Manufacturing, Agribusiness, and Services Projects

Supporting factors. Expansion meant that a project company benefited from expansion and market consolidation through acquisitions or organic growth, driving cost efficiencies and economies of scale. It could also mean that they had higher capital expenses or larger project scope than expected because of more investment. Relationship management meant that the project was a repeat deal with the same sponsors and gained from previous experience, or the International Finance Corporation had an active portfolio management and was flexible by helping the clients address their pressing needs in a more depressed market environment. Business factor as a favorable factor meant that the project gained from the increased market opportunity or its business model provided an edge over the competition.

(continued)

Box G.1. Examples of Supporting and Constraining Factors Affecting IFC Investment Project Performance, by Industry Group (cont.)

Constraining factors. Environment and sustainability meant that the project had material shortcomings in meeting environmental and social requirements or that the client did not have in place some of the required important corporate policies. Business factors meant that the project company had shortcomings in the business model or it suffered from unfavorable business and operating environment or industry cyclicality.

Disruptive Technologies and Funds Projects

Supporting factors. Technical expertise and track record as a positive factor meant fund managers with strong capacity or relevant experience. The environment and sustainability as a favorable factor meant the investment fund had high environmental and social and corporate governance standards.

Constraining factors. Project size meant that the fund was unable to reach its target size. This could be due to the fund manager's lack of experience or mean that the fund's investment thesis was too risky. Technical expertise and track record as an adverse factor typically meant that the fund manager lacked experience in private equity investing, in the specific fund target segment, or in emerging markets. Customers typically meant that the fund deviated from its investment strategy and invested in different types of portfolio companies than intended at approval. For example, the fund may have invested in developed countries rather than in emerging markets.

Source: Independent Evaluation Group.

Note: IFC = International Finance Corporation.

Investment projects with low development outcomes were affected by the same adverse factors as the entire portfolio but with a larger impact. The top two factors affecting their performance were (i) unfavorable economic issues, which affected 35 percent of projects; and (ii) high business risks, which affected 29 percent of projects. Another critical constraining factor affecting poor performing projects was limited technical expertise and track record of sponsors and clients, which affected 24 percent of projects.

Moreover, many weak-performing investment projects were challenged by increased competition (20 percent) and the effects of epidemics and the COVID-19 (14 percent).

A similar pattern was observed among investment projects in the Africa region, which had relatively low development outcome ratings. The performance of investment projects in this region was challenged mostly by (i) adverse economic factors (29 percent), (ii) limited technical expertise and track record of sponsors and clients (21 percent), and (iii) high business risks (17 percent). These three factors were also among the top factors that negatively affected the performance of investment projects in fragile and conflict-affected situation (FCS) countries. The factors above adversely affected a similar share of FCS projects, except for business risks, which affected a larger share (27 percent). The performance of investment projects in FCS countries was also hampered by low asset quality (20 percent), increased legal and regulatory risks (20 percent), and issues of low integrity and transparency (20 percent).

The performance factors of investment projects with borderline development outcome ratings of mostly unsuccessful or mostly successful were different. The borderline projects rated mostly unsuccessful were hampered by adverse country and market factors such as (i) unfavorable economic issues (41 percent of projects), (ii) high business risks (31 percent), (iii) increased competition (28 percent), and (iv) epidemics and COVID-19 (18 percent). The macroeconomic factors and the pandemic were unforeseen and beyond the control of IFC and sponsors, while factors such as business risks and competition could potentially be mitigated. Moreover, these challenges could not be compensated for by the limited supporting factors present in these projects. With regard to their performance, only a small share of mostly unsuccessful investment projects benefited from strong technical expertise and track record of sponsors and clients (13 percent); high capacity, capitalization and leverage of sponsors (10 percent); strong environmental and social performance (5 percent); and high integrity and transparency (5 percent).

In contrast, borderline investment projects rated mostly successful for development outcome have been negatively affected by country and market factors such as (i) epidemic and COVID-19 (22 percent), (ii) unfavorable economic issues (17

percent), (iii) foreign exchange issues (17 percent), and (iv) project-specific environment and sustainability issues (17 percent). However, these were countered with (i) strong technical expertise and track record of sponsors and clients (43 percent), (ii) conducive business aspects (13 percent), (iii) aligned commitment and motivation of sponsors (9 percent), and (iv) favorable technology choices, which allowed these investment projects to adequately cope with the challenges and achieve largely positive results. All these factors could be considered within the control of IFC and the sponsors. Table G.1 presents examples of potential measures that could be taken by IFC to mitigate adverse performance factors.

Table G.1. Examples of Potential Mitigation Measures for IFC Investment Projects, by Industry Group

Adverse Performance Factor	Potential Mitigation Measures
Financial Institutions Group Projects	
Business factors: Refers to a financial institution moving away from lending to the target beneficiaries (for example, micro, small, and medium enterprises; agribusiness; affordable housing; and so on) because of a more adverse environment, which will reduce development impact.	Mitigant: Provide technical assistance to the financial institution either before or during the implementation to enhance its capacity to increase or maintain lending to risky but highly developmental segments.
Integrity, transparency, fairness, and reputation: Refers to internal integrity due diligence issues with the sponsor affecting the project.	Mitigant: Conduct ongoing integrity due diligence to ensure that issues do not materialize. Proper supervision could help project teams react quickly to try to mitigate any adverse integrity due diligence issues during project implementation in a timely manner.
Loan factors: Refers to the entire facility not being drawn down or disbursed due to a change in the financial institution's strategy. It also refers to the loan tenor not being appropriate for the project or loan covenants not being appropriate or followed.	Mitigant: Carefully assess the financial institution's strategy and capacity at appraisal to ensure commitment to the project's development objectives (for example, micro, small, and medium enterprise lending). Ensure that the loan is properly priced or appropriate for the financial institution's purposes.
Infrastructure and Natural Resources Projects	
Business factors: Refers to flawed, untested, or fragile business models or a slowdown in market growth.	Mitigant: Assess the viability of the business model during appraisal. Decline to invest when the business model is flawed. Provide additionality to assist the client in improving operations and practices.

(continued)

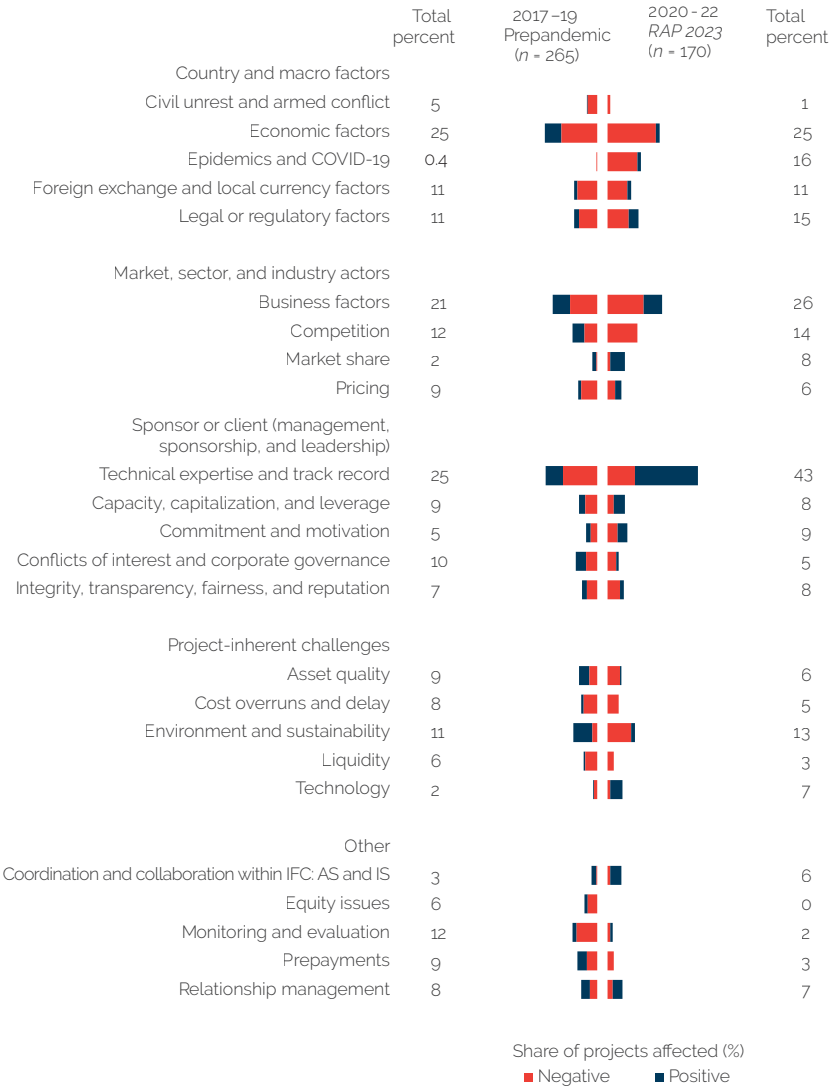
Adverse Performance Factor	Potential Mitigation Measures
Technical expertise and track record: Refers to the sponsors not measuring up to what the project was aiming to achieve without adequate operational and financial capacity, depth of management, and relevant experience.	Mitigant: Closely examine the sponsor's financial capacity, management depth, and relevant experience at appraisal. Provide additionality and active portfolio supervision if sponsor decides to invest.
Manufacturing, Agribusiness, and Services Projects	
Environment and sustainability: Refers to the project having material shortcomings in meeting environmental and social requirements or the client not putting in place required corporate policies.	Mitigant: In case of corporate financing investments, ensure than the client has in place all required corporate policies and that all its businesses comply with environmental and social performance standards. The cost of environmental and social improvements needs to be estimated at appraisal and included in the project cost if needed. Provide additionality to improve client's environmental and social practices.
Business factors: Refers to shortcomings in the business model or unfavorable business and operating environment or industry cyclicality.	Mitigant: Assess the viability of the business model during appraisal. Provide additionality to clients in improving their operations and practices.
Disruptive Technologies and Funds Projects	
Project size: Refers to a fund not reaching its target size, potentially due to fund manager's lack of experience or riskiness of investment thesis.	Mitigant: Provide additionality by assisting fund manager in fundraising. Conversely, decline to invest in the fund if the fund is unable to reach the minimum capital.
Technical expertise and track record: Typically refers to a fund manager lacking experience in private equity investing, in the specific fund target segment, or in emerging markets.	Mitigant: Provide additionality through technical assistance to both the fund manager and the downstream portfolio companies to help make the fund successful.
Customers: Typically refers to fund deviation from its investment strategy and investing in different types of portfolio companies than intended at approval.	Mitigant: Through position on the advisory committee, voice objections to any unnecessary deviations in the fund strategy or decline to participate in investments that are not in line with the investment thesis as presented at approval.

Source: Independent Evaluation Group.

Note: IFC = International Finance Corporation.

The key factors affecting CY20–22 investment project performance were broadly the same as those affecting CY17–19 project performance. However, the CY20–22 investment projects were more negatively affected by country- and market-related factors than CY17–19 investment projects not exposed to the pandemic (see figure G.1).

Figure G.1. Factors Affecting IFC Investment Project Performance: The Prepandemic Cohort Compared with the *RAP 2023* Cohort



Source: Independent Evaluation Group.

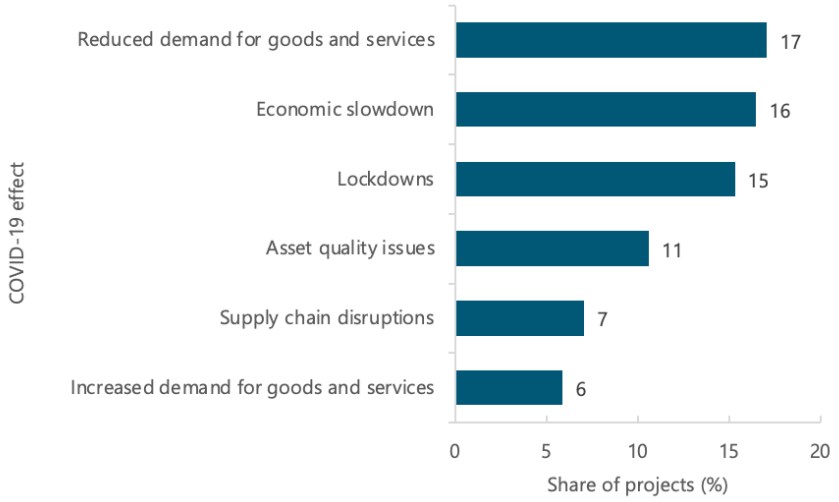
Note: The factor identification for calendar year 2020–22 projects was based on human thinking, whereas for calendar year 2017–19 prepandemic projects, it was based on machine learning. Positive = the identified factor aided the project performance. Negative = the identified factor constrained the project performance. AS = advisory services; IFC = International Finance Corporation; IS = investment services; *RAP = Results and Performance of the World Bank Group*.

Compared with the prepandemic investment projects, a larger proportion of CY20–22 investment projects was negatively affected by country and market factors such as (i) adverse economic issues, (ii) high business risks, (iii) unforeseen epidemics and COVID-19, and (iv) high competition. In addition, the environment and sustainability factor was also more prevalent for CY20–22 investment projects. In contrast, monitoring and evaluation issues and unfavorable market pricing were prevalent negative factors for the prepandemic investment projects.

These findings suggested that CY20–22 investment projects operated in more challenging country and market conditions than did CY17–19 projects. They were more negatively affected by adverse macroeconomic factors, unforeseen epidemics and COVID-19, high business risks, and high competition, while also dealing with low ability and expertise of sponsors and clients. However, there were some supporting factors that aided the performance of CY20–22 investment projects, of which the most prevalent was the strong technical expertise and track record of sponsors and clients. This factor was particularly present in high-performing investment projects, enabling them to effectively cope with challenges posed by the pandemic.

To glean more insights, *RAP 2023* further explored how the pandemic affected the investment projects and how the portfolio adapted to the operating conditions posed in that context. The pandemic presented various unforeseen challenges for investment project implementation and performance, often aggravating already existing issues in projects and depressing performance. According to project evaluation documents, the pandemic affected over half of investment projects through reduced demand for goods and services, economic slowdown, lockdowns, asset quality issues, and supply chain disruptions (see figure G.2). Some investment projects, especially investment funds with investees in several sectors, were affected by the pandemic in multiple ways. In a few cases, the pandemic aided the investment project's operations, by increasing the demand for goods and services produced by the client companies. This happened particularly in the health care, retail, and manufacturing sectors, which supplied goods and services essential to cope with COVID-19.

Figure G.2. Main Ways COVID-19 Pandemic Affected Investment Project Implementation



Source: Independent Evaluation Group.

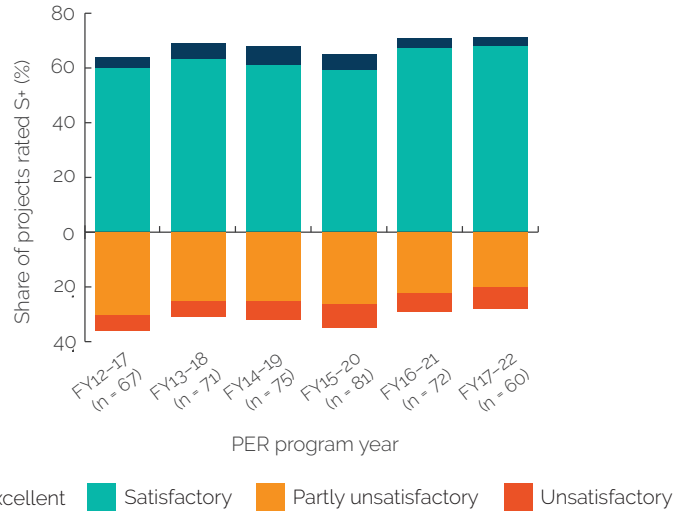
While the pandemic initially presented significant challenges for the private sector in emerging markets, many companies showed remarkable resilience and adaptability. In the financial sector, most IFC clients contracted their loan portfolio and focused on asset quality issues. Many real sector project companies implemented cost-saving initiatives to increase efficiency and shore up margins. Others invested quickly in information technology solutions to facilitate remote work. Many companies rolled out online versions of their business lines, particularly companies in the higher education and food and consumer retail sectors. In a few cases, the pandemic increased the demand for clients' goods and services. For example, in the health care sector, investment project companies began producing COVID-19 tests and vaccines, while project hospitals began treating COVID-19–affected patients. The private sector reacted quickly to the changing economic landscape, demonstrating its adaptability, resilience, and flexibility in a challenging environment.

Appendix H. MIGA Guarantee Project Rating Trends and Patterns

This appendix presents trends and patterns of Multilateral Investment Guarantee Agency (MIGA) guarantee project performance across various project categories, such as sector, region, country lending group, and country fragile and conflict-affected situation (FCS) status, as of June 30, 2023.

Development outcome success ratings of MIGA guarantee projects remained stable on a six-year rolling basis, with 72 percent of projects in fiscal years (FY)17–22 rated satisfactory or better (S+), a similar level to FY16–21 (71 percent; figure H.1). It should be noted, however, that MIGA's overall development outcome success rates were lower in the last three years, partially reflecting the more challenging operating environment. A larger share of negatively rated projects in FY20–22 offset higher-performing projects in FY17–19 (figure H.2). The composition of granular development outcome ratings of MIGA guarantee projects in FY17–22 remained largely the same as in FY16–21.

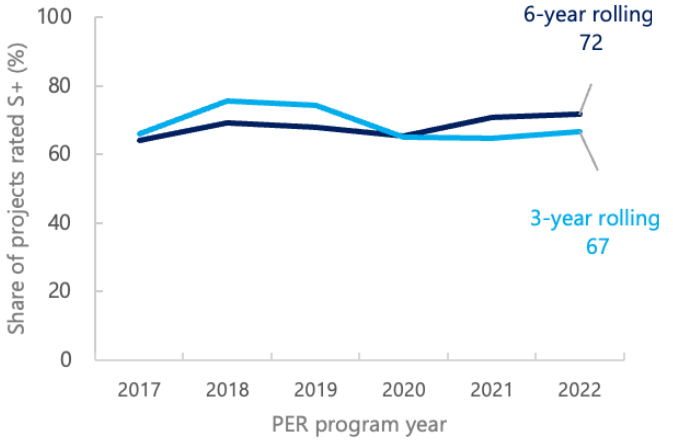
Figure H.1. MIGA Guarantee Project Development Outcome Ratings



Source: Independent Evaluation Group PER database.

Note: The MIGA Project Evaluation Report guidelines were changed in FY19, replacing a four-point scale for development outcome ratings with a six-point scale. The six-point rating scale, applied to projects starting in fiscal year 2020, was converted to a four-point one as follows: highly successful = excellent, successful and mostly successful = satisfactory, mostly unsuccessful = partly unsatisfactory, and highly unsuccessful and unsuccessful = unsatisfactory. FY = fiscal year; MIGA = Multilateral Investment Guarantee Agency; PER = Project Evaluation Report; S+ = satisfactory or better.

Figure H.2. Development Outcome Ratings (6-year rolling versus 3-year rolling)



Source: Independent Evaluation Group PER database.

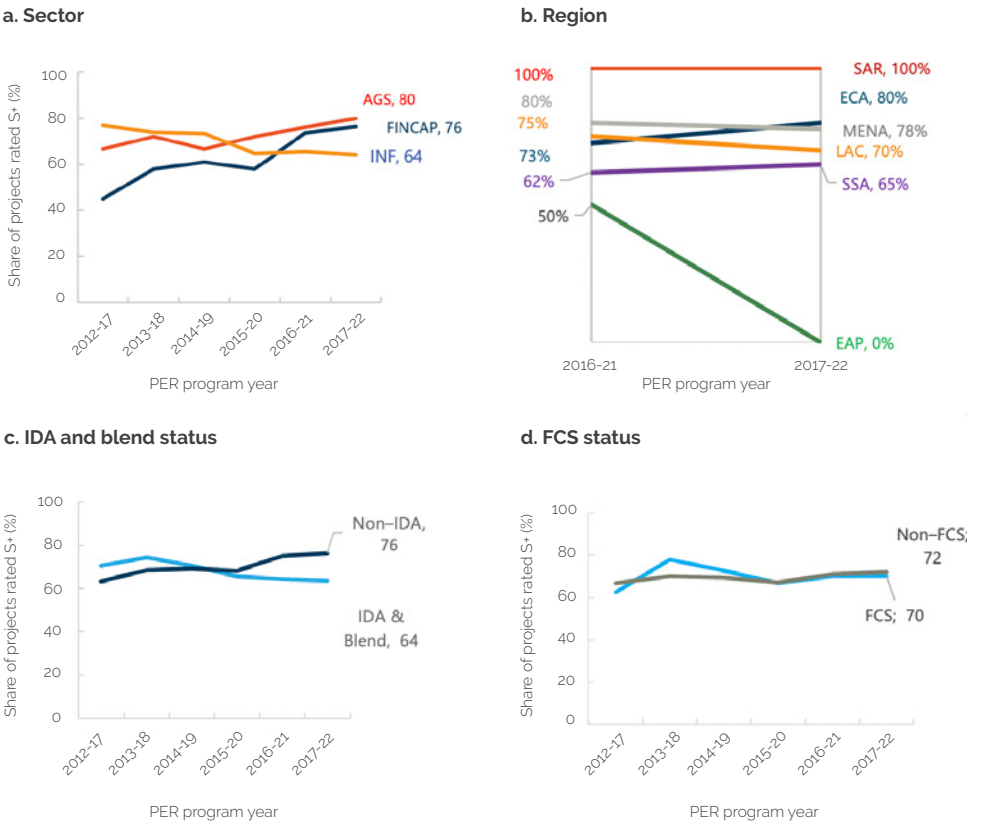
Note: PER = Project Evaluation Report.

At the sector level, the guarantee projects in Finance and Capital Markets, and Agribusiness and General Services have steadily improved their development outcome performance, while the Infrastructure and the Energy and Extractive Industries sectors combined projects had a slight decline. For Finance and Capital Markets projects, which represented 28 percent of MIGA's evaluated portfolio, development outcome success ratings increased from 74 percent in FY16–21 to 76 percent in FY17–22 (figure H.3, panel a). The Agribusiness and General Services guarantee projects, accounting for a quarter of MIGA's evaluated portfolio, improved their success ratings from 76 percent to 80 percent over the same period. In contrast, the Infrastructure and the Energy and Extractive Industries sectors combined projects, which comprised the largest share of MIGA's evaluated portfolio, had a slight decline in their development outcome success rates from 66 percent to 64 percent in FY17–22, from the previous period. However, this decline in the average ratings of Infrastructure sector guarantee projects did not have a sizable effect on MIGA's overall development outcome success rates.

At the regional level, no significant rating declines or improvements have been observed across regions. Given the small number of evaluated guarantee projects per region, the results must be interpreted with caution. Having

said that, guarantee projects in the South Asia (note small number of projects), Europe and Central Asia, Middle East and North Africa (note small number of projects), and Latin America and the Caribbean (note small number of projects) regions generally had higher development outcome success ratings than those in other regions. In contrast, the lower performance of guarantee projects in the Sub-Saharan Africa and East Asia and Pacific (note small number of projects) regions were dragging down MIGA's overall success rate in FY17–22 (figure H.3, panel b).

Figure H.3. MIGA Guarantee Projects Rated Satisfactory or Better for Select Project Categories



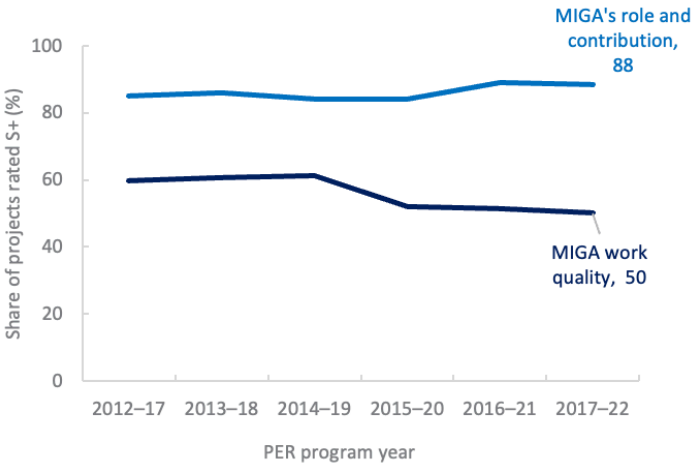
Source: Independent Evaluation Group.

Note: AGS = Agribusiness and General Services; EAP = East Asia and Pacific; ECA = Europe and Central Asia; FCS = fragile and conflict-affected situation; FINCAP = Finance and Capital Markets; IDA = International Development Association; INF = Infrastructure and the Energy and Extractive Industries; LAC = Latin America and the Caribbean; MENA = Middle East and North Africa; PER = Project Evaluation Report; S+ = satisfactory or better; SAR = South Asia; SSA = Sub-Saharan Africa.

The performance gap between guarantee projects in International Development Association (IDA) and blend countries and those in non-IDA countries continued to persist during FY17–22. MIGA's overall development outcome success ratings in guarantee projects in both IDA and blend and non-IDA countries remained stable. However, the performance of IDA and blend projects, representing 37 percent of MIGA's evaluated portfolio, continued to be lower than that of non-IDA projects, with 64 percent of those rated S+ for the overall development outcome in FY17–22. In contrast, the success rate for guarantee projects in non-IDA countries, representing 63 percent of the MIGA's evaluated portfolio, remained at 76 percent in FY17–22 (figure H.3 panel c). This is worth attention because one of MIGA's priority areas has been on facilitating foreign direct investment in IDA and blend countries. MIGA's overall development outcome performance in guarantee projects in FCS countries was on par with that of non-FCS projects. Although the evaluated portfolio of MIGA guarantee projects in FCS countries was generally small, with 10 or fewer projects per period, the percentage of those rated S+ for overall development outcome remained stable at 70 percent in FY17–22 compared with FY16–21 (figure H.3, panel d).

MIGA achieved high success rates in delivering its expected role and contribution. The percentage of guarantee projects with MIGA's role and contribution rated satisfactory was 88 percent in FY17–22, the same as in FY16–21 (figure H.4). MIGA's role and contribution was most significant in environmental and social areas and reducing some project risks. MIGA achieved its expected role and contribution in 80 percent of guarantee projects in FCS countries. MIGA's role and contribution success ratings of guarantee projects in IDA and blend countries were on par with those in non-IDA countries.

Figure H.4. MIGA Work Quality and Role and Contribution Success Ratings



Source: Independent Evaluation Group PER database.

Note: MIGA = Multilateral Investment Guarantee Agency; PER = Project Evaluation Report; S+ = satisfactory or better.

Half of guarantee projects in FY17–22 had shortcomings in MIGA work quality. MIGA work quality success ratings were S+ in only 50 percent of MIGA guarantee projects in FY17–22, which was similar to 51 percent in FY16–21, but still lower than 60 percent in FY12–17 (see figure H.4). During FY20–22, only 40 percent of projects had S+ MIGA work quality ratings, which was lower than 48 percent in FY19–21 and 56 percent in FY15–17. MIGA work quality was correlated with development outcomes of 75 percent of all guarantee projects in FY17–22. Twenty-seven percent of guarantee projects had low development outcomes, which were associated with weak work quality.

Appendix I. Analysis of Multilateral Investment Guarantee Agency Guarantee Project Outcome Types

The goal of the outcome type analysis is to describe the intended development outcomes of Multilateral Investment Guarantee Agency (MIGA) guarantee projects and assess the relationship between those outcomes and performance ratings. This analysis added value because MIGA guarantee projects are not rated at the outcome level. The analysis expands on the outcome analysis conducted in the 2021 *Results and Performance of the World Bank Group* (*RAP*) report. For this analysis, the Independent Evaluation Group conducted a deep dive into 16 MIGA guarantee projects evaluated and validated from fiscal years (FY)20–22. It should be noted that the results should be interpreted cautiously given the small number of MIGA guarantee projects evaluated by MIGA and validated by the Independent Evaluation Group.

Multilateral Investment Guarantee Agency Outcome Types

MIGA's guarantee projects were focused on higher-level outcomes as envisaged by their specific intended outcomes. There were 16 MIGA guarantee projects in the *RAP* cohort from FY20–22, which contained 78 outcomes, of which 62 were project-level outcomes and 16 were foreign investment–level outcomes. There were, on average, 5 outcomes per project, consisting of 4 project-level outcomes and 1 foreign investment–level outcome. MIGA's outcome types reflected its focus on larger infrastructure projects, such as transportation, energy, telecom, and health care for the FY20–22 period. Overall, the *RAP* identified 30 outcome types (25 project level and 5 foreign investment level) for MIGA guarantee projects that were based on the Impact Measurement and Project Assessment Comparison Tool (IMPACT) system (table I.1).

As shown in table I.1, all MIGA guarantee projects pursued the project-level outcome of improved access to goods and services. Other prevalent project-level outcomes included increased employment and quality and affordability of goods

and services, which were present in 63 percent of all projects. Among foreign investment–level outcomes, market development was the most common outcome type, present in 56 percent of projects. Business and sector practices was the only other prevalent foreign investment–level outcome.

Table I.1. MIGA Guarantee Projects by All Outcome Types: Prevalence

Outcome Type	FY20–22 (*n* = 16)	
	Outcomes (no.)	Share of projects (%)
Project-level outcomes		
1.1 – Access to goods and services	19	100
1.1.1 – Access to goods and services (MSME)	1	6
1.1.2 – Access to goods and services (female)	0	0
1.1.3 – Access to goods and services (customers)	4	25
1.1.4 - Access to goods and services (miscellaneous)	14	75
1.2 - Quality and affordability of goods and services	10	50
1.3 - Enhanced capacity of final beneficiaries	0	0
1.4 - Improved living standards (earnings) of individuals	2	13
1.5 - Improved sales and profitability of enterprises	0	0
1.6 – Economic return	1	6
1.7 - Financial and business performance of direct clients	0	0
2.1 - Suppliers and distributors reached	0	0
2.2 - Improved capacity of suppliers and distributors	0	0
2.3 - Improved sales and profitability of suppliers and distributors	0	0
3.1 - Increased employment	10	63
3.2 - Improved capacity and skills	1	6
3.3 - Improved earning of employees	0	0
4.1 - Increased transfers to the government	6	38

(continued)

Outcome Type	FY20–22 (*n* = 16)	
	Outcomes (no.)	Share of projects (%)
Project-level outcomes		
5.1 - Increased money spent and transfer to the communities	0	0
6.1 - Enhanced E&S standards of the client	0	0
6.2 - Greenhouse gas reduction	6	38
6.3 - Efficient use of resources	3	19
7.1 - Gross value added	0	0
7.2 - Induced and indirect employment	1	6
7.3 - Export sales	2	13
8.1 - Governance	1	6
Total project-level outcomes	62	
Foreign investment–level outcomes		
9 - Business and sector practices	4	25
10 - Market development	9	56
11 - Development reach	1	6
12 - Sustainability	0	0
13 - Signaling effects	2	13
Total foreign investment–level outcomes	16	

Note: Individual guarantee projects can have multiple outcome types. E&S = environmental and social; MIGA = Multilateral Investment Guarantee Agency; MSME = micro, small, and medium enterprise.

Prevalence. The prevalence of outcome types in MIGA guarantee projects changed across the *RAP 2021* and *RAP 2023* portfolios, but the results should be interpreted cautiously given a small number of evaluated and validated MIGA guarantee projects. Furthermore, the sample sizes are very different. Of note, the prevalence of the project-level outcome of access to goods and services increased from 85 percent to 100 percent. Market development increased in prevalence in MIGA guarantee projects since *RAP 2021* (table I.2).

Table I.2. Comparison of Prevalence of MIGA Guarantee Project Outcomes

Outcome Type	FY17–19 (*RAP 2021*; *n* = 39)		FY20–22 (*RAP 2023*; *n* = 16)	
	Outcomes (no.)	Share of projects (%)	Outcomes (no.)	Share of projects (%)
Project-level outcomes (top 10)				
1.1 - Access to goods and services	54	85	19	100
1.1.4 - Access to goods and services (miscellaneous)	n.a.	n.a.	14	75
1.1.3 - Access to goods and services (customers)	46	77	4	25
1.1.1 - Access to goods and services (MSME)	9	18	1	6
3.1 - Increased employment	21	54	10	63
1.2 - Quality and affordability of goods and services	29	51	10	50
1.2.1 - Quality of goods and services	n.a.	n.a.	6	38
1.2.2 - Affordability of goods and services	n.a.	n.a.	4	13
4.1 - Increased transfers to the government	18	44	6	38
6.2 - Greenhouse gas reduction	7	18	6	38
6.3 – Efficient use of resources	7	13	3	19
1.4 - Improved living standards (earnings) of individuals	2	5	2	13

(continued)

Outcome Type	FY17–19 (*RAP 2021*; *n* = 39)		FY20–22 (*RAP 2023*; *n* = 16)	
	Outcomes (no.)	Share of projects (%)	Outcomes (no.)	Share of projects (%)
7.3 - Export sales	4	8	2	13
1.6 - Economic return	10	23	1	6
3.2 - Improved capacity and skills	12	28	1	6
7.2 - Induced or indirect employment	6	13	1	6
8.1 - Governance	4	8	1	6
Foreign investment–level outcomes				
10 - Market development	24	41	9	56
9 - Business and sector practices	15	33	4	25
13 - Signaling effects	1	3	2	13
11 - Development reach	0	0	1	6
12 - Sustainability	0	0	0	0

Source: Independent Evaluation Group.

Note: MIGA = Multilateral Investment Guarantee Agency; MSME = micro, small, and medium enterprise.

Achievement. During FY20–22, MIGA projects had a higher probability of achieving project-level outcomes than foreign investment–level outcomes. Of the total of 78 outcomes, 50 percent were fully achieved and 22 percent were partially achieved. Of the 62 project-level outcomes, 55 percent were fully achieved and 21 percent were partially achieved (table I.3). MIGA guarantee projects also had 16 foreign investment–level outcomes, of which only 31 percent were fully achieved and 25 percent were partially achieved. The outcome achievement rates confirmed the *RAP*'s hypothesis that foreign investment–level outcomes would be more challenging to achieve.

Among prevalent project-level outcomes, quality and affordability of goods and services had the highest achievement rate at 80 percent. Increased employment was second with a 60 percent achievement rate. Among foreign investment–level outcomes, signaling effects had the highest achievement rate at 100 percent, while market development, the most prevalent foreign investment–level outcome, only had a 33 percent achievement rate.

Shortcomings in project monitoring and evaluation affected the validation of some intended outcomes. Of the 78 total outcomes, 24 percent were not achieved, including 24 percent of project-level outcomes and 44 percent of foreign investment–level outcomes. A significant percent of outcomes were not tracked by MIGA through the Development Effectiveness Indicator System (DEIS) monitoring system. Of the total outcomes, 69 percent were not tracked in DEIS, meaning that only 31 percent of MIGA guarantee project outcomes were tracked. Of the outcomes not tracked in DEIS, 72 percent were project-level outcomes and 28 percent were foreign investment–level outcomes. In such cases, either MIGA or the Independent Evaluation Group used other information sources to validate the outcome claim. However, 10 percent of total outcomes could not be verified because of a lack of evidence and were coded as "cannot be verified." Thirteen percent of foreign investment–level outcomes (compared with 10 percent of project-level outcomes) could not be verified because of a lack of evidence, confirming the hypothesis that foreign investment–level outcomes were more challenging to measure than project-level outcomes.

Table I.3. MIGA Outcome Type Performance: Achievement Rate

Outcome Type	Outcomes (no.)	Outcome Achieved (fully; %)	Outcome Achieved (partially; %)	Outcome Achieved (fully + partially; %)
Project-level outcomes (top 10)				
1.1 - Access to goods and services	19	63	11	74
1.1.1 - Access to goods and services (MSME)	1	0	0	0
1.1.2 - Access to goods and services (female)	0	n.a.	n.a.	n.a.
1.1.3 - Access to goods and services (customers)	4	100	0	100
1.1.4 - Access to goods and services (miscellaneous)	14	57	14	71
1.2 - Quality and affordability of goods and services	10	80	10	90
1.4 - Improved living standards (earnings) of individuals	2	0	50	50
1.6 - Economic return	1	0	0	0
3.1 - Increased employment	10	60	40	100
3.2 - Improved capacity and skills	1	100	0	100
4.1 - Increased transfers to the government	6	17	67	84
6.2 - Greenhouse gas reduction	6	33	0	33

(continued)

Outcome Type	Outcomes (no.)	Outcome Achieved (fully; %)	Outcome Achieved (partially; %)	Outcome Achieved (fully + partially; %)
Project-level outcomes (top 10)				
6.3 - Efficient use of resources	3	67	33	100
7.2 - Induced or indirect employment	1	100	0	100
7.3 - Export sales	2	50	0	50
8.1 - Governance	1	0	0	0
Total project-level outcomes	62	55	21	76
Foreign investment–level outcomes				
9 - Business and sector practices	4	0	0	0
10 - Market development	9	33	44	77
11 - Development reach	1	0	0	0
12 - Sustainability	0	n.a.	n.a.	n.a.
13 - Signaling effects	2	100	0	100
Total foreign investment–level outcomes	16	31	25	56
Total outcomes	78	50	22	72

Source: Independent Evaluation Group.

Note: Of the total project-level outcomes, 10 percent were considered not achieved because the results could not be verified. Of the total foreign investment–level outcomes, 13 percent were considered not achieved because their results could not be verified. MIGA = Multilateral Investment Guarantee Agency; MSME = micro, small, and medium enterprise; n.a. = not applicable.

As shown in table I.4, outcome achievement rates changed in different ways between the *RAP 2021* and *RAP 2023* cohorts. However, these results should be interpreted cautiously given the small number of evaluated and validated MIGA guarantee projects. Furthermore, the sample size for *RAP 2023* (n = 16) was much smaller than that for *RAP 2021* (n = 39), which could create statistical differences. This is borne out by the fact that only the achievement rates for quality and affordability of goods and services and access to goods and services (customers) were statistically significant between the two *RAP* cohorts.

Table I.4. Comparison of MIGA Outcome Achievement Rates

Outcome Type	FY17–19 (*RAP 2021*; n = 39)		FY20–22 (*RAP 2023*; n = 16)	
	Outcomes (no.)	Outcomes achieved (%)	Outcomes (no.)	Outcomes achieved (%)
Project-level outcomes (top 10)				
1.1 - Access to goods and services	54	56	19	63
1.1.4 - Access to goods and services (miscellaneous)	n.a.	n.a.	14	57
1.1.3 - Access to goods and services (customers)*	46	57	4	100
1.1.1 - Access to goods and services (MSME)	9	56	1	0
3.1 - Increased employment	21	38	10	60
1.2 - Quality and affordability of goods and services**	29	52	10	80
1.2.1 - Quality of goods and services	n.a.	n.a.	6	83
1.2.2 - Affordability of goods and services	n.a.	n.a.	4	75
4.1 - Increased transfers to the government	18	33	6	17

(continued)

Outcome Type	FY17–19 (*RAP 2021*; *n* = 39)		FY20–22 (*RAP 2023*; *n* = 16)	
	Outcomes (no.)	Outcomes achieved (%)	Outcomes (no.)	Outcomes achieved (%)
Project-level outcomes (top 10)				
6.2 - Greenhouse gas reduction	7	57	6	33
6.3 – Efficient use of resources	7	71	3	67
1.4 - Improved living standards (earnings) of individuals	2	100	2	0
7.3 - Export sales	4	100	2	50
1.6 - Economic return	10	60	1	0
3.2 - Improved capacity and skills	12	33	1	100
7.2 - Induced or indirect employment	6	50	1	100
8.1 - Governance	4	0	1	0
Foreign investment–level outcomes				
10 - Market development	24	29	9	33
9 - Business and sector practices	15	40	4	0
13 - Signaling effects	1	100	2	100
11 - Development reach	0	n.a.	1	0
12 - Sustainability	0	n.a.	0	n.a.

Source: Independent Evaluation Group.

Note: MIGA = Multilateral Investment Guarantee Agency; MSME = micro, small, and medium enterprise; n.a. = not applicable.
*Statistically significant at 90% C.I. **Differences statistically significant at 95% C.I.

Multilateral Investment Guarantee Agency Outcome Ratings Analysis

As with International Finance Corporation investment projects, *RAP 2023* conducted an outcome analysis to answer evaluation question 2, "What has been the evolution of development outcomes pursued, measured, and achieved at the project level, and what is the relationship between outcomes and project performance ratings?" As per the Project Evaluation Report guidelines, the development outcome ratings of MIGA guarantee projects are assigned at the project level and subdimension level and not at the specific outcome level. This analysis added value because it examined the relationship between underlying outcomes and development outcome ratings.

The outcome analysis showed that MIGA guarantee projects that achieved more of their intended outcomes and foreign investment–level outcomes also achieved higher development outcome ratings. As shown in table I.5, there is a clear link between outcome achievement rates and development outcome ratings. Projects with a successful development outcome rating achieved 80 percent of their outcomes, while those with an unsuccessful rating achieved only 22 percent of their outcomes. However, achievement of foreign investment–level outcomes was not as much of a determining factor, but this could be explained by the very small sample size.

Nevertheless, the analysis showed that successful guarantee projects achieved 89 percent of their project-level outcomes and 50 percent of their foreign investment–level outcomes, whereas unsuccessful guarantee projects achieved only 25 percent of their project-level outcomes and none of their foreign investment–level outcomes. Financial and environmental and social performance also played a role in explaining development outcome ratings but not as much as outcome achievement. Outcome achievement was also the main difference between "borderline" projects (mostly successful to mostly unsuccessful): 69 percent versus 45 percent. The higher the level of outcome achievement, the higher the development outcome rating, and vice versa.

Table I.5. MIGA Development Outcome Ratings and Outcome Achievement Rates

Development Outcome Rating	Projects (no.)	Outcomes (no.)	Overall Outcome Weighted Achievement Rate (%)	Project-Level Outcome Weighted Achievement Rate (%)	Foreign Investment–Level Outcome Weighted Achievement Rate (%)	Project Business Performance Average Rating	Environmental and Social Effects Average Rating
Successful	5	25	80	89	50	3.0	3.0
Mostly successful	5	24	69	79	43	2.4	3.2
Mostly unsuccessful	4	20	45	44	50	2.0	2.8
Unsuccessful	2	9	22	25	0	1.0	3.0

Source: Independent Evaluation Group.

Note: Achievement rates at the project level are measured by efficacy of outcome achievements in projects, with the following weights: outcome achieved = 1, partly achieved = 0.5, not achieved = 0, and cannot verify = 0. Project business performance and environmental and social effects ratings' numerical values are as follows: excellent = 4, satisfactory = 3, partly unsatisfactory = 2, and unsatisfactory = 1. MIGA = Multilateral Investment Guarantee Agency.

Appendix J. Analysis of Factors Affecting MIGA Guarantee Project Implementation and Performance

The goal of this appendix is to describe the factors of success or failure that affected Multilateral Investment Guarantee Agency (MIGA) guarantee project implementation and performance in the COVID-19 pandemic context and uncover patterns across projects and country characteristics based on a qualitative review.

All MIGA guarantee projects in the combined fiscal years (FY)20–22 Project Evaluation Report program covered in the deep-dive analysis were evaluated and validated after March 2020. The implementation of half of MIGA guarantee projects were affected by COVID-19 at least to some extent, particularly given that 81 percent of them were still active in MIGA portfolio at the time of evaluation. On average, the guarantee projects in the 2023 *Results and Performance of the World Bank Group* (*RAP*) cohort were exposed to COVID-19 for 27 percent of their active project lives. During the pandemic, there were some delays in the delivery of MIGA self-evaluations, which to some extent influenced the profile of guarantee projects analyzed in the *RAP* cohort.

Factors Affecting the Project Implementation and Performance in the COVID-19 Context

This *RAP* conducted a deep-dive analysis of 16 MIGA guarantee projects validated from FY20–22 in which it identified the top three critical factors explaining the performance for each project. The *RAP* leveraged the existing taxonomy of performance factors developed by the Independent Evaluation Group and consisting of 5 categories and 51 subcategories (see appendix A for definition of different factors). Each factor can have either a positive or negative influence on the project performance. Among the 16 projects, 10 had high development outcome ratings and 6 had low development outcome ratings.

The top factors that negatively affected FY20–22 projects were (i) unfavorable foreign exchange factors (25 percent), (ii) cost overruns and construction delays (25 percent), (iii) increased competition (19 percent), and (iv) inadequate monitoring and evaluation (19 percent). The projects were also challenged by (i) adverse economic issues (13 percent), (ii) epidemics and the COVID-19 (13 percent), (iii) increased political risk (13 percent), (iv) low technical expertise, and track record (13 percent), and (v) inadequate market assessment (13 percent). Economic issues, epidemics and COVID-19, and political risks factors were not foreseeable and could be considered outside of the control of MIGA, the sponsor, or the project enterprise. However, other factors could be considered within their control and thus could have been mitigated. Among the top adverse factors, foreign exchange issues affected Agribusiness and General Services projects, while competition, cost overruns and construction delays and monitoring and evaluation factors were more prevalent in Infrastructure projects.

The most prevalent factors supporting the performance of FY20–22 MIGA guarantee projects were (i) high technical expertise and track record of sponsors and clients (31 percent), (ii) favorable legal and regulatory aspects (31 percent), (iii) high market share of the clients' business (13 percent), (iv) favorable business aspects (13 percent), (v) positive environment and sustainability aspects (13 percent), and (vi) savings in project cost and construction times (13 percent). Among these factors, legal and regulatory factors could be considered outside of the control of MIGA, sponsors, and project enterprises, but the other factors could be considered within their control. Technical expertise and track record and legal and regulatory factors featured more prominently in four Agribusiness and General Services projects, while project cost and construction times was more relevant for Infrastructure projects. Box J.1 illustrates examples on how factors affected performance of guarantee projects in various sectors.

Box J.1. Examples of Supporting and Constraining Factors Affecting Performance of MIGA Guarantee Projects, by Sector

Finance and Capital Market Projects

Supporting factors. Positive technical expertise and track record meant that the financial institution management was experienced and the financial institution had a strong performance in terms of total loan growth, earnings, capital adequacy, and liquidity. Market share meant that the financial institution was the leader in the small and medium enterprise lending market, which gave it an edge over the competition.

Constraining factors. Weak earnings and profitability meant that the financial institution had weaker-than-expected loan growth and financial performance, including deteriorated asset quality, lower profitability, and tighter liquidity. Inadequate market assessment meant that MIGA did not adequately assess the financial institution's commitment and capacity to achieve the project's intended objectives, given that the ultimate beneficiaries of the MIGA-supported financing were not realized as expected, thus reducing the development impact.

Infrastructure, and Energy and Extractive Industries Projects

Supporting factors. Positive project cost and construction times meant that the project construction was under budget and that implementation was on time.

Constraining factors. Adverse competition meant that the project enterprise experienced a highly competitive market that depressed revenues, suffered from overcapacity in the sector, or faced competition from more efficient new entrants in the market. Cost overruns and construction delays meant that the project suffered from implementation delays and cost overruns, which negatively affected the project's financial and economic returns. Monitoring and evaluation issues meant that the project had shortcomings in monitoring and evaluation, such as the lack of quantified baseline or targets and information on actual results, preventing the verification of achievement of its intended development impacts.

(continued)

Box J.1. Examples of Supporting and Constraining Factors Affecting Performance of MIGA Guarantee Projects, by Sector (cont.)

Agribusiness and General Services Projects

Supporting factors. The high technical expertise and track record meant that the sponsors had financial capacity, relevant experience in implementing public-private partnership projects, and competent management that ensured high-quality operations and maintenance. The positive legal and regulatory factors meant that the payment mechanism in public-private partnership projects protected the sponsors from the downside risk of a depreciation of the local currency.

Constraining factors. The adverse foreign exchange and local currency factors meant that depreciation of the local currency and resulting foreign exchange losses negatively affected financial results of the project enterprises.

Source: Independent Evaluation Group.

Note: MIGA = Multilateral Investment Guarantee Agency.

In low-performing projects, the key factors were related to market, project, sponsor, and other issues. The critical factors constraining the performance of six low-performing projects were (i) higher competition (50 percent), (ii) cost overruns and construction delays (50 percent), (iii) low ability, technical expertise and track record of sponsors and clients (33 percent), and (iv) inadequate market assessment (33 percent). Competition, cost overruns and construction delays, and inadequate market assessment also hampered the performance of the four borderline projects rated mostly unsuccessful for development outcome. All these factors could be considered to have been within the control of MIGA, sponsors, and the project enterprise, and therefore could have been mitigated.

In contrast, the five borderline projects rated mostly successful for development outcome were negatively affected by (i) adverse economic issues (40 percent), (ii) foreign exchange factors (40 percent), and (iii) monitoring and evaluation issues (40 percent). Economic issues and foreign exchange factors were beyond MIGA, sponsor, and project enterprise's control, while monitoring

and evaluation was within their control. The important difference between borderline projects was that the adverse factors affecting mostly successful projects were countered with the strong ability, technical expertise and track record of sponsors and project enterprises, and favorable legal and regulatory aspects. These mitigants enabled these borderline projects to adequately cope with the challenges of the pandemic and multiple crises and achieve largely positive results. Table J.1 presents examples of potential measures that MIGA could use to mitigate adverse performance factors.

Table J.1. Examples of Potential Mitigation Measures for MIGA Guarantee Projects, by Sector

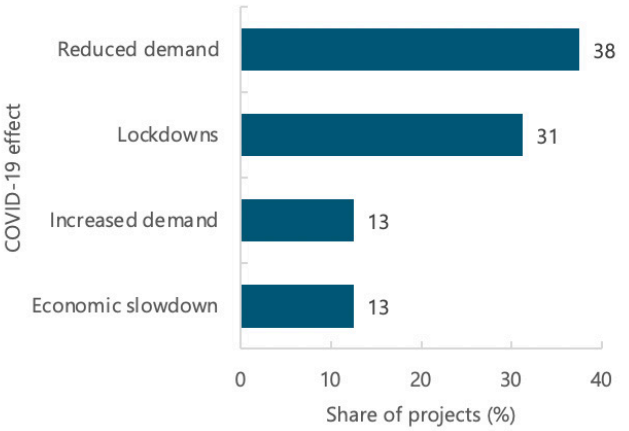
Adverse Performance Factor	Potential Mitigation Measures
Finance and Capital Market Projects	
Inadequate market assessment: Refers to the inadequate assessment of a financial institution's commitment and capacity to achieve the project's intended objectives. In such cases, the beneficiaries of the financing are not supported as expected, reducing development impact.	Mitigant: Better define development impact objectives at appraisal by clarifying the purpose and use of proceeds of guaranteed facilities and establishing appropriate development impact indicators.
Infrastructure, and Energy and Extractive Industries Projects	
Monitoring and evaluation: Refers to the project having shortcomings in monitoring and evaluation, such as the lack of quantified baseline or targets and information on actual results, preventing the verification of achievement of its intended development impacts.	Mitigant: Establish appropriate mechanisms for development impact data gathering in guarantee projects, where the project enterprise is not a direct signatory to supported financing agreements.
Agribusiness and General Services Projects	
Foreign exchange and local currency: Refers to depreciation of the local currency that results in foreign exchange losses that negatively affect the financial results of the project enterprise.	Mitigant: Identify and assess the potential impact of foreseeable macroeconomic developments, including depreciation of the local currency, that may increase the size of the government's financial obligations and assess whether the government will be willing and have the capacity to pay the increased obligations.

Source: Independent Evaluation Group.

Note: MIGA = Multilateral Investment Guarantee Agency.

RAP 2023 explored further the main ways the pandemic affected the MIGA guarantee project implementation and how the portfolio adapted to the changing conditions. The COVID-19 pandemic affected MIGA guarantee projects through lockdowns and economic slowdowns (see figure J.1). Some projects were affected by the pandemic in multiple ways. Lockdowns particularly affected the infrastructure projects in the public transportation sector, reducing the demand for their services. Lockdowns also affected some hospital projects in the Agribusiness and General Services sector by reducing demand for elective medical services and procedures considered nonessential during the pandemic. Conversely, for other hospital projects that operated as designated COVID-19 treatment hospitals, the pandemic supported the project's operations by increasing the demand for health care services targeting the COVID-19 patients.

Figure J.1. Main Ways COVID-19 Affected MIGA Guarantee Project Implementation



Source: Independent Evaluation Group.

Information in project evaluation documents about how the MIGA guarantee project enterprises adapted to the challenging environment in the pandemic context was limited, but the most apparent cases of adaptation were related to hospital projects, which added capacity for high-quality intensive care and hospital beds and helped the government respond to the medical demands the pandemic created.

IEG
Independent Evaluation Group

WORLD BANK GROUP
World Bank • IFC • MIGA

The World Bank
1818 H Street NW
Washington, DC 20433