Approach Paper

World Bank Group Scorecard Formative Evaluation

November 3, 2025

1. Background and Context

1.1 The World Bank first adopted a Corporate Scorecard in 2011.[1] Using a similar format to the International Development Association's 2002 and subsequent results frameworks, it comprised two tiers of indicators tracking development results and two tracking elements of Bank performance. Its scope was expanded to include the International Finance Corporation (IFC) and the Multilateral Investment Guarantee Agency (MIGA) in 2014, and the four tiers reduced to three with the combination of operational and organizational performance into one category. There were further iterations in 2017 and 2020, each of which presented similar indicators under somewhat varying headings. The 2023 version is structured around 8 vision indicators, 22 client context indicators, 22 World Bank Group results indicators, 15 results narratives, and an organizational dashboard (World Bank Group 2025; appendix C).

1.2 Compared with its predecessors, the 2023 Scorecard is intended to measure, at the apex of the institution, the Bank Group's contribution to development outcomes rather than simply tracking the outputs it delivers (table 1.1), with a focus on indicators that capture people's well-being (World Bank 2024; World Bank Group 2025). It is also intended to better account for the full range of the Bank Group's contribution to development outcomes through results narratives that describe how it contributes to change beyond its financing, through knowledge transfer and capacity building. By harmonizing indicator definitions and flagging results achieved through joint engagements, it is expected to incentivize cooperation across the World Bank, IFC, and MIGA. In addition, the Scorecard aims to reinstate the Bank Group's leadership among other multilateral development banks (MDBs) in defining outcome measures for emerging priorities such as digital development.

1.3 The 2023 redesign of the Scorecard took place in the context of the Bank Group's Evolution. The Evolution process introduced a "new playbook," with reforms to deliver impact at scale (World Bank 2023a). The key priorities included leveraging knowledge, strengthening the country engagement cycle, and scaling up private sector activities and private capital mobilization. They were to be supported by changes in the World Bank's operating and financial models, in particular more effective collaboration across the Bank Group and better alignment of incentives and management practices with outcomes. The phased release of Scorecard indicators in FY 2024–25 was framed as a milestone in this process, intended to motivate the adoption of the reforms and monitor

their results. Management intends to cascade the Scorecard's outcome orientation into country engagement and portfolio design.

Table 1.1. Building Blocks of the 2023 Scorecard

	Vision		Outcome Areas	
Building Blocks		**Client context**	**World Bank Group results**	**Results narratives**
What do they do?	Trace the global goals to which the World Bank Group expects to contribute	Describe the development context in which the Bank Group operates	Track outcomes from Bank Group interventions	Narrate Bank Group efforts not captured by indicators
What for?	Set the North Star for the Bank Group's ambition	Reality check for Bank Group progress toward results	Yardstick to assess Bank Group development effectiveness	Capture the effects of policy and institutional reforms and other investments at the country level
Bank Group attribution	Not attributable to Bank Group	Line of sight to Bank Group mission and vision	Attributable to Bank Group interventions	Plausible contribution

Source: World Bank 2024.

1.4 The Independent Evaluation Group (IEG) has not previously evaluated the Scorecards as a stand-alone topic. However, its report on the Bank Group's self-evaluation systems included Corporate Scorecards in its scope (World Bank 2016). The report found that these tools had some advantages, providing an easy-to-understand perspective on the performance of the institution and enabling aggregation across heterogeneous areas of engagement, but it also noted inherent limitations. Aggregate indicators oversimplify the reality of complex portfolios and can be skewed by results in a few large countries. Imposing common metrics crowds out more context-specific metrics that are necessary for adaptive management.

1.5 Other IEG evaluations have also assessed how the Bank Group manages for results. They have found repeatedly that the system is geared more toward external reporting and upward accountability than organizational learning. By focusing excessively on project ratings and achieving corporate targets, it limits candor and disincentivizes high-quality self-evaluation (World Bank 2016, 2020). The evaluations have also drawn attention to the tendency for management to reward new lending and rapid disbursement at the expense of longer-term commitments (World Bank 2024). IEG's evaluation of the capital increase package discussed accountability in the context of reforms requested by shareholders and pointed out that corporate indicators were insufficient to capture the World Bank's contribution to development outcomes (World Bank 2023b).

2. Objectives and Intended Users of the Evaluation

2.1 The objective of this evaluation is to assess the fit for purpose of the Bank Group Scorecard during its development and rollout—as a tool for improving the measurement of outcomes, management for outcomes, accountability for outcomes to the Board of Executive Directors, and communication of outcomes to external audiences. It is also formative, intended to identify areas for improvement and influence ongoing adjustments. It will inform the midterm management review of the 2023 Scorecard's implementation during FY26 and is being undertaken in tandem with a Group Internal Audit report.[2]

2.2 The evaluation's primary audience is the members of the Committee on Development Effectiveness, Bank Group senior management, the World Bank's Outcomes Department, and IFC's and MIGA's Scorecard teams.

3. Preliminary Theory of Change

3.1 Public sector scorecards originated as an attempt to improve the effectiveness and legitimacy of government. As a legacy of "new public management" and "results-based management" theories, they reflect an acknowledgment that public institutions must meet several objectives. These are to ensure that social outcomes are clearly defined and delivered, that taxpayer or donor resources are being used efficiently in pursuit of these outcomes, and that the beneficiaries of the service and those paying for it perceive it to be effective and in accordance with their values (Barber 2017). In theory, scorecards work by drawing attention to these objectives and informing internal and external audiences as to whether they are being met. By defining aspirational goals, even if these are subsequently unmet, they can induce effort and progress beyond what would have been achieved in their absence. Like other results-based management tools, scorecards are also intended to narrow the asymmetry of information between an organization's board and its management (Raimondo 2018). They can be useful tools for agenda setting by board members and an entry point to strategy discussions that would otherwise be dominated by management perspectives (Drew and Kaye 2007). In some cases, their intention has also been to free managers from up-front controls and reduce compliance costs. This autonomy in how objectives are achieved can be valuable in contingent and idiosyncratic settings requiring adaptation to context; it can also deter implementers from blaming externally imposed constraints for an absence of progress (Perakis and Savedoff 2015). Yet the effectiveness of scorecards depends very much on how they are designed and used.

Design

3.2 Because scorecards aim to improve effectiveness and accountability by prioritizing some activities and objectives over others, they should be focused on an organization's most important goals or those around which there is a strategic consensus. Indicators should capture what is important, not just what is quantifiable. They should also measure something that those responsible for meeting them can reasonably be expected to influence, with targets set and adjusted prudently, reflecting an understanding of past trajectories, some judgment as to future performance, and a willingness to explain why adjustments might be warranted. The data reported should be reliable, unambiguous, and verifiable by those to whom individuals and the organization are accountable. The indicators chosen must capture well what they intend to measure. They should also be comparable across reporting periods and collectible at reasonable cost (UK National Audit Office 2001). Scorecards can be valuable in communicating an organization's vision to outsiders, but such statements need to be plausible and to be backed with precise, durable objectives and a persuasive theory of change (Lowcock and Dissanayake 2024). The supporting blend of narratives and data should be tailored to the audiences to whom they are addressed—whether donors, recipients, professionals, or the public (OECD 2021). They should not downplay the complexity and riskiness of aid spending or overestimate the contribution of those providing it (Valters and Whitty 2017).

Use

3.3 Scorecards are only one part of a performance management system. They may provide information as to whether an organization is managing its inputs well and achieving the goals it has set for itself, but that information must be used judiciously. To be effective, an organization must be capable not only of generating performance data but also of interpreting and responding to it through its planning and programming, budgeting, and human resource systems. Scorecards work best when the relationship between inputs and outcomes is known and stable, and a metric is a reasonable proxy for outcomes. The more complex and diverse the outcomes an organization seeks to achieve are (or the more it needs to tailor its activities to circumstances), the more it will need to invest in strong evaluation systems and rely on judgment and deliberation in its decision-making (Bridges and Woolcock 2024; Honig 2018; World Bank 2020). Otherwise, it risks distorting behavior by focusing attention on wasteful or counterproductive activities (Dahler-Larsen 2019; Raimondo 2018). Poor implementation can lead to a perception of increased workload and reduce the time available to employees for reflective learning (Antonsen 2014). Experience indicates that centralized performance management systems can be gamed by those whose behavior they seek to influence (Hood 2006). They have also been criticized for inhibiting a focus on "real-world

delivery," undermining client ownership, disincentivizing risk-taking, and failing to measure unintended outcomes, whether positive or negative (Lowcock and Dissanayake 2024; Valters and Whitty 2017). There is an additional risk of their being used primarily as tools for seeking legitimacy, rather than as mechanisms for improving organizational effectiveness and performance (Chang 2007).

3.4 Using scorecards as an accountability tool also implies more than the provision of information. Accountability differs from transparency in that it requires not just reporting but the capacity, or the right, to demand answers and the capacity to sanction if those answers are not forthcoming or are unsatisfactory (Schedler 1999). It depends, therefore, on the behavior not only of those being held to account but also of the overseers. Board oversight, in the sense of policing performance, is more likely to be effective when (i) the board's duties are clearly distinguishable from those of management such that the latter cannot claim co-responsibility for failure, (ii) the board or some other outside authority approves performance benchmarks, and (iii) there is strong consensus among shareholders on what those benchmarks should be (Martinez-Diaz 2009).

3.5 Effective external communication requires more than just the dissemination of data and narratives. Public bodies need to have an awareness of what taxpayers and users think about their programs, as well as knowledge of what influences those perceptions and a strategy for improving them. This includes a willingness to respond to misinformation and disinformation (OECD 2021).

Implicit World Bank Group Scorecard Theory of Change

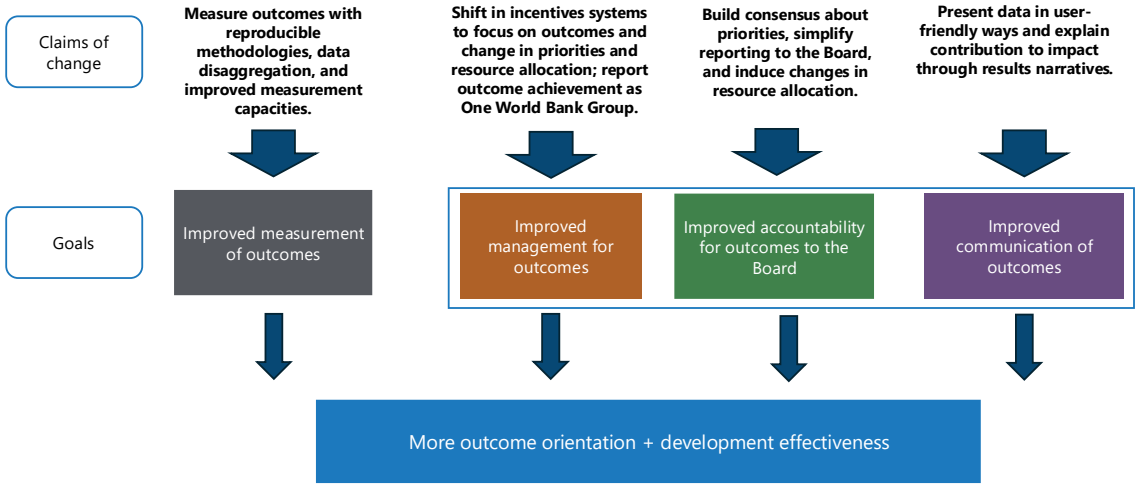
3.6 The purpose of the Bank Group's 2023 Scorecard is to increase the institution's outcome orientation and development effectiveness. To achieve that goal, the Scorecard expects to improve the measurement of outcomes, management for outcomes, accountability of the institution for outcomes to the Board, and communication of outcomes to external audiences (World Bank 2024; figure 3.1).

Measurement of Outcomes

3.7 The Scorecard expects to (i) improve the measurement of outcomes through reproducible and transparent methodologies that focus on the beneficiary-level impact, (ii) aggregate and report results from projects, and (iii) disaggregate results by demographic groups and client segments (World Bank 2024). It also uses results narratives to capture long-term support for outcomes that combines financing, technical assistance, and knowledge transfer.

3.8 For those claims to be true, the Scorecard relies on several assumptions. These include that the Bank Group results indicators be truly outcome level, that the expected and achieved values of those results indicators be calculated accurately from project data and consistently across the three main institutions comprising the Bank Group, that data disaggregation is accurate, and that the results narratives reflect a true theory of the Bank Group's contribution.

Figure 3.1. Implicit Scorecard Theory of Change



Source: Independent Evaluation Group.

Management for Outcomes

3.9 The Scorecard claims to improve how the World Bank manages for results by drawing attention to and rewarding the achievement of outcomes rather than outputs and activities. The selection of indicators is intended to help the organization prioritize and deepen areas of engagement. Comparing the achievement of Bank Group results indicators with a subset of those indicators with 2030 targets and client context indicators allows resources to be allocated either to where they are having the most impact or to where they are most needed,[3] while organizational effectiveness indicators seek to measure whether inputs are being used efficiently (World Bank 2024). By integrating reporting across the Bank Group and making clear which results are due to the joint or individual efforts of Bank Group institutions, the Scorecard is also expected to motivate an increase in joint engagements.

3.10 These claims assume that Scorecard indicators are used to drive incentives (for example, in staff performance management discussions and institutional awards) and that they influence the scope of country programs and project portfolios, even if these are jointly determined by other corporate priorities and market demand. They also

assume that Bank Group results indicators at least contribute to improvements in client context indicators, in both nature and scale; that targets reflect a plausible trajectory of change; and that organizational effectiveness indicators are associated with the achievement of Bank Group results.

Accountability to the Board

3.11 The Scorecard also claims to improve accountability for outcomes to the Board, by building consensus around a subset of development objectives and by simplifying and integrating reporting of outcomes across the Bank Group.

3.12 Those claims assume that executive directors agree on common priorities, that they use the Scorecard to set those priorities and to evaluate whether they are being met, and that they can demand an explanation from management if they are not.

External Communications

3.13 The Scorecard claims to improve the communication of outcomes through dashboard tools, including interactive visualizations. It also uses results narratives to explain the contribution of the Bank Group beyond lending operations at the country or sector level and over a longer period than individual projects. In addition, it uses the Bank Group's convening power to increase harmonization in reporting across MDBs.

3.14 These claims assume that Scorecard information reaches its intended audiences, that the results narratives are credible, and that harmonization of indicators across MDBs is a desirable and shared goal.[4]

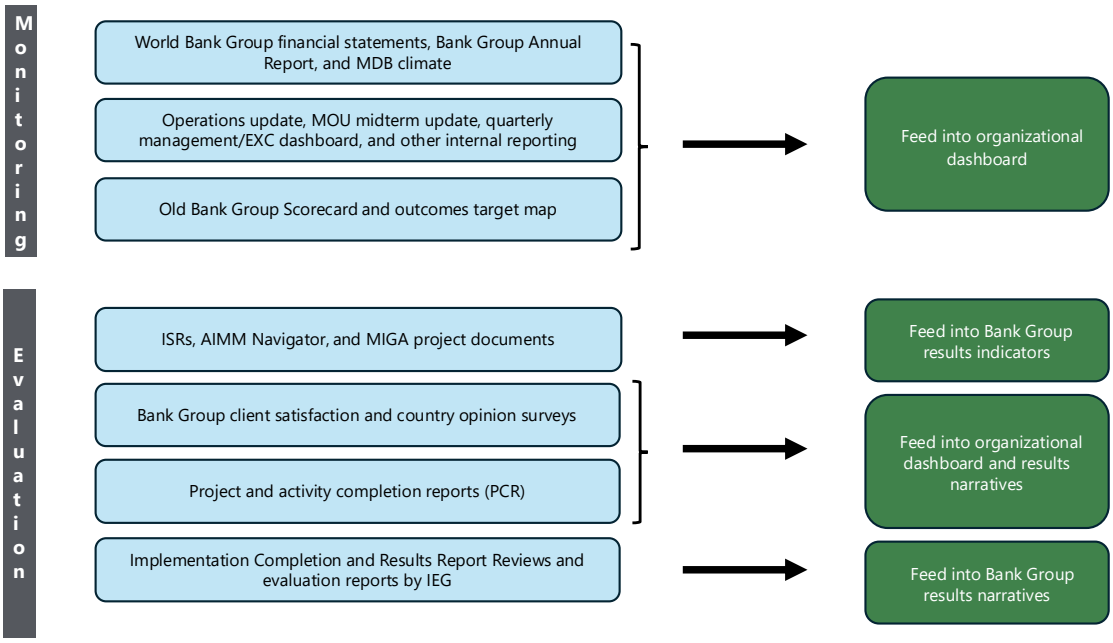
3.15 The Scorecard theory of change relies on several elements present in other performance management systems—the main one being its use of indicators to increase the salience of meaningful objectives and link incentives to them. However, it differs from some in continuing, through organizational effectiveness indicators, to regulate all parts of the project delivery chain, including inputs and outputs. The Bank Group also differs from many public organizations in that its projects are designed with and implemented by client governments, and most of the time in the case of IFC and MIGA by the private sector. The market-driven nature of IFC and MIGA operations and restrictions on the publication of confidential private client data complicate target realism and transparency.

Relationship Between the Scorecard and World Bank Group Monitoring and Evaluation Systems

3.16 The 2023 Scorecard largely resembles previous versions in its fit within the Bank Group monitoring and evaluation system. The Bank Group results indicators are

sourced from implementation status reports and in some cases project documents for World Bank operations, mainly through the Anticipated Impact Measurement and Monitoring Navigator for IFC investment operations, and directly from individual project reports for MIGA operations. Dashboard indicators use data from (i) internal reporting systems, in particular the World Bank memorandum of understanding midterm update and the World Bank and IFC operations updates, and (ii) internal and external evaluations, in particular IEG operations and monitoring and evaluation quality ratings, and client satisfaction surveys. The results narratives—a new feature in the 2023 version—draw from (i) evaluative sources, such as impact evaluations, client surveys, and IEG evaluations, and (ii) operational, research, or communications sources, such as client official websites and documents, stakeholder testimonies, Bank Group communication documents (blogs and briefs), and media and journal articles (figure 3.2).

Figure 3.2. World Bank Group Scorecard and Monitoring and Evaluation Systems



Source: Independent Evaluation Group.
Note: AIMM = Anticipated Impact Measurement and Monitoring; EXC = Office of the President; IEG = Independent Evaluation Group; ISR = Implementation Status and Results Report; MDB = multilateral development bank; MIGA = Multilateral Investment Guarantee Agency; MOU = memorandum of understanding; PCR = Project Completion Report.

4. Evaluation Scope and Questions

4.1 The formative evaluation will focus on the design and use of the 2023 Scorecard. First, it will examine whether the methodologies on which the 2023 version is based are sound and whether, compared with previous versions, it has improved the outcome orientation of Bank Group results indicators, the relevance of the organizational

effectiveness indicators and the balance between them and Bank Group results indicators, and the clarity and persuasiveness with which Scorecard information is communicated. Second, it will assess whether the implementation of the 2023 Scorecard followed a clear plan with adequate support, whether it has had any influence on the quality and accuracy of the data reported, whether its priorities have been reflected in country programs and operational design, whether it has altered the behavior of the Board or the use of Scorecard data by clients and other external audiences, and whether it is contributing to the harmonization of results reporting across MDBs.

4.2 The evaluation will also examine the Scorecard's coherence with other parts of the Bank Group's results-based management system, in particular the country engagement model and the Bank Group–wide monitoring systems, including the country and project-based rating systems. It will evaluate the use of Scorecard information in setting performance incentives and influencing staff behavior. While it may identify tensions between the Scorecard and those systems, including any unintended consequences of the Scorecard itself, it will not assess the effectiveness of recent reforms to the Country Partnership Framework or the self-evaluation systems on which some of the indicators are based. Nor will it undertake a cost-benefit analysis of the Scorecard or examine the performance of the Outcomes Department.

4.3 The temporal scope of the evaluation starts from the first quarter of FY24, when a preliminary version of the Scorecard approach was presented to the Board. The closing period for all evaluation questions (EQs) is the fourth quarter of FY25.

4.4 The EQs are as follows:

1. To what extent has the 2023 Bank Group Scorecard improved the measurement of outcomes within the institution?

2. To what extent do the design, implementation, and use of the 2023 Bank Group Scorecard serve its intended purposes of improving management for outcomes, accountability to the Board, and external communication about outcomes?

3. Are there any risks or unintended negative consequences in the design or use of the Scorecard? What mitigation measures or complementary approaches might be put in place to address them?

4.5 The evaluation design matrix in appendix A lists the subquestions.

5. Evaluation Design

Design Principles

5.1 This is a formative evaluation, the purpose of which is to assess the early implementation of the 2023 Scorecard and provide lessons to inform its evolution. It will therefore involve frequent consultations with the Scorecard's designers, implementers, and users to ensure that its findings are useful. IEG proposes to establish an Evaluation Reference Group comprising representatives of these stakeholder groups, which will meet at least three times during the preparation, implementation, and reporting phases of the evaluation. The Evaluation Reference Group will serve as a sounding board for the IEG team, helping ensure that the evaluation reflects stakeholder perspectives and is factually accurate, useful, and timely. It will not interfere with the independence of the evaluators or influence findings. The IEG team will retain final editorial authority over the report. The evaluation is also theory based, in that it will assess the validity of the conditions on which the Scorecard's theory of change rests. We will triangulate evidence from multiple stakeholders and methods (for example, indicator analysis, interviews, and content analysis, among others) to reach valid inferences that are systematically informed by the evidence.

Building Blocks

5.2 The evaluation will be supported by the following building blocks executed in the following sequence. The detailed design is explained in appendix A.

1. Structured literature review of prior evaluations and research on the design, use, and effectiveness, including unintended consequences, of scorecards in other MDBs and international organizations (EQ 1 and EQ 3).

2. Bank Group results indicator analysis: review of Bank Group results indicator methodologies, data sets from FY24 and FY25, and metadata to assess whether indicators truly measure what they set out to measure, whether they are indeed outcomes, and the extent to which they capture expected Bank Group contributions. It will also assess the construction or plausibility of expected, achieved values and targets, and the accuracy of data disaggregation. The analysis will examine a sample of underlying indicator project data to evaluate the reliability and completeness of the data used (EQ 1).

3. Bank Group results narrative analysis: review of methodologies and results narratives from FY24 and FY25 to assess whether they draw on credible evaluative evidence, adequately capture Bank Group nonlending or other indirect contributions to long-term outcomes, and provide usable information and lessons (EQ 1 and EQ 2).

4. Content analysis of (i) Country Partnership Frameworks, Performance and Learning Reviews, country portfolios, and project documents from FY24 to FY25 to assess the extent to which the Scorecard has led to prioritization of Scorecard objectives or other changes in country engagement (EQ 2.2); (ii) Board discussions and statements by Board members to assess the extent to which they use Scorecard information in discussions of country strategies (EQ 2.3); and (iii) analysis of the Scorecard's external webpages to analyze who accesses the information and for what purpose (EQ 2.4).

5. Analysis of interviews and workshops with (i) the Scorecard's designers, implementers, and users (see appendix B for the interview sampling frame) to understand the challenges faced during rollout and implementation, including unintended consequences (EQ 2.1 and EQ 3), and (ii) similar stakeholders in other MDBs and international organizations to assess the extent to which the 2023 Bank Group Scorecard has influenced results reporting across MDBs (EQ 2.4). The team will consult with the executive directors with respect to the use of Scorecard information by clients but not directly with either public or private sector clients.

6. Analysis of surveys of World Bank task team leaders and IFC investment officers and advisory team leaders to understand the changes they have faced after the introduction of the Scorecard (for example, changing Country Partnership Framework priorities, changes in results frameworks, and retrofitting of Bank Group results indicators to approved projects; EQ 1, EQ 2.1, and EQ 2.2).

7. Quantitative analysis to assess the historical correlation between (i) selected dashboard indicators and Bank Group results indicators and (ii) the contribution of Bank Group results indicators to client context indicators (EQ 2.2).

5.3 Appendix B presents the sampling frame and proposed sampling strategies for each of the relevant stakeholders.

Risks and Mitigation

5.4 The main limitation of the evaluation is that the evaluand is a single case. We will, however, draw on evidence from the experience of other institutions and, where

possible, from previous Bank Group Scorecards. Given the high-level attention to the Corporate Scorecard from the Board, World Bank management, and external stakeholders, IEG will maintain regular communication with World Bank management and the Committee on Development Effectiveness during evaluation design and implementation, including sharing emerging findings once available.

6. Quality Assurance Process

6.1 We will follow IEG's quality assurance processes to ensure the rigor and usefulness of the evaluation's design, analyses, and findings. Processes include internal review processes, data collection protocols for all components, and review by the IEG methods team. Three independent peer reviewers will also review the evaluation design and methodologies and the final report: Karin Finkelston (former vice president, Partnerships, Communication, and Outreach, IFC); José Guilherme Reis (former executive director for Brazil, Inter-American Development Bank); and Mark Singleton (consultant, Independent Evaluation Group).

7. Resources

Timeline

7.1 The evaluation draft and IEG management review are planned for the third quarter of FY26. Bank Group management review and e-submission are planned for the fourth quarter of FY26. The main output will be a report.

Budget

7.2 The budget for delivering the evaluation is US$675,000, of which US$66,000 was incurred in FY25.

Team

7.3 The evaluation team will be led by Thomas Kenyon (senior economist and task team leader). Team members include Santiago Tellez Cañas (evaluation officer; methods support); Arunjana Das (extended-term consultant; communications analysis and structured literature review); Diana Goldemberg (evaluation officer; quantitative analysis, management, and accountability analysis); Luke Jordan (consultant; econometric analysis); Olivia Podmore (analyst; methods analysis); Estelle Raimondo (head, Methods; methods adviser and overall guidance); and Ichiro Toda (senior evaluation officer; adviser on IFC and MIGA).

[1] IFC started to use its own scorecard in the 1990s as part of its annual strategy discussion with the Board. Since the approval of the capital increase package in 2018, it has continued to use a revamped version covering five buckets of indicators: development impact (ex ante and ex post), financial sustainability, program and client delivery, efficiency, and diversity and inclusion.

[2] The Group Internal Audit is an independent advisory body that assesses whether the risk management, control, and governance processes of World Bank entities are adequately designed and functioning effectively. Its report will focus on the adequacy of process design and implementation. Specifically, it will assess whether (i) processes for data collection and aggregation have been defined and implemented; (ii) Scorecard data have been verified for accuracy, completeness, and timeliness; (iii) roles, responsibilities, and decision-making authority have been established for preparing and reporting the Scorecard; (iv) staff members have been provided with training and support in implementing the Scorecard; and (v) there are secure and reliable systems for the consistent and accurate production of the Scorecard. IEG and the Group Internal Audit will collaborate to avoid overburdening management with requests for information and to ensure the consistency of their findings.

[3] These include beneficiaries of social safety net programs; people receiving quality health, nutrition, and population services; people provided with access to electricity; women using broadband access; and women-led businesses using financial services.

[4] Harmonization across MDBs might in theory reduce the reporting burden on clients and enable comparisons of development effectiveness. It assumes, however, that all such institutions focus on the same outcome and share the same ability to report on them.

Bibliography

ADB (Asian Development Bank). 2019. *Corporate Results Frameworks and Scorecards and the Experience of Selected Multilateral Development Bank and Bilateral Organizations*. ADB.

Antonsen, Yngve. 2014. "The Downside of the Balanced Scorecard: A Case Study from Norway." *Scandinavian Journal of Management* 30 (1): 40–50.

Barber, Michael. 2017. *Delivering Better Outcomes for Citizens: Practical Steps for Unlocking Public Value.* HM Treasury.

Bridges, Kate, and Michael Woolcock. 2022. "Measuring What Matters: Principles for a Balanced Data Suite That Prioritizes Problem-Solving and Learning." Policy Research Working Paper 10051, World Bank.

Cha, Seungman. 2017. "The Impact of the Worldwide Millenium Development Goals Campaign on Maternal and Under-Five Child Mortality Reduction: 'Where Did the Worldwide Campaign Work Most Effectively?'" *Global Health Action* 10 (1): 1267961.

Chang, Li-cheng. 2007. "The NHS Performance Assessment Framework as a Balanced Scorecard Approach: Limitations and Implications." *International Journal of Public Sector Management* 20 (2): 101–17.

Dahler-Larsen, Peter. 2019. "The Skeptical Turn in Evaluation." In *The Evaluation Enterprise: A Critical View*, edited by Jan-Eric Furubo and Nicoletta Stame, 58–80. Routledge.

Drew, Stephen A., and Roland Kaye. 2007. "Engaging Boards in Corporate Direction Setting: Strategic Scorecards." *European Management Journal* 25 (5): 359–69.

Honig, Dan. 2018. *Navigation by Judgement*. Oxford University Press.

Honig, Dan. 2019. "When Reporting Undermines Performance: The Costs of Politically Constrained Organizational Autonomy in Foreign Aid Implementation." *International Organization* 73 (1): 171–201.

Hood, Christopher. 2006. "Gaming in Targetworld: The Targets Approach to Managing British Public Services." *Public Administration Review* 66 (4): 515–21.

Lowcock, Mark, and Ranil Dissanayake. 2024. *The Rise and Fall of the Department for International Development*. Center for Global Development.

Martinez-Diaz, Leonardo. 2009. "Boards of Directors in International Organizations: A Framework for Understanding the Dilemmas of Institutional Design." *The Review of International Organizations* 4: 383–406.

Milner, Helen V., and Dustin Tingley. 2013. "Public Opinion and Foreign Aid: A Review Essay." *International Interactions* 39 (3): 389–401.

OECD (Organisation for Economic Co-operation and Development). 2021. *OECD Report on Public Communication: The Global Context and the Way Forward*. OECD.

Perakis, Rita, and William Savedoff. 2015. "Does Results-Based Aid Change Anything? Pecuniary Interests, Attention, Accountability and Discretion in Four Case Studies." Policy Paper 052, Center for Global Development.

Raimondo, Estelle. 2018. "The Power and Dysfunctions of Evaluation Systems in International Organizations." *Evaluation* 24 (1): 26–41.

Schedler, Andreas. 1999. "Conceptualizing Accountability." In *The Self-Restraining State: Power and Accountability in New Democracies*, edited by Andreas Schedler, Larry Diamond, and Marc F. Plattner. Lynne Rienner Publishers.

UK National Audit Office. 2001. *Choosing the Right Fabric: A Framework for Performance Information.* National Audit Office.

Valters, Craig, and Brendan Whitty. 2017. *The Politics of the Results Agenda in DFID.* Overseas Development Institute.

World Bank. 2016. *Behind the Mirror: A Report on the Self-Evaluation Systems of the World Bank.* Independent Evaluation Group. World Bank.

World Bank. 2020. *The World Bank Group Outcome Orientation at the Country Level.* Independent Evaluation Group. World Bank.

World Bank. 2023a. *Evolution of the World Bank Group: A Report to Governors*. World Bank.

World Bank. 2023b. *The World Bank Group's 2018 Capital Increase Package: An Independent Validation of Implementation and Results*. Independent Evaluation Group. World Bank.

World Bank. 2024. *New World Bank Group Scorecard FY24–FY30: Driving Action, Measuring Results.* World Bank.

World Bank. 2025. "World Bank Group Scorecard Board Update." World Bank.

World Bank Group. 2025. "The World Bank Group Scorecard." World Bank Group. https://scorecard.worldbank.org/en/home.

Appendix A. Evaluation Design Matrix

Table A.1 indicates the key questions, sources, and data collection and analysis methods, along with the strengths and limitations associated with these.

Table A.1. Evaluation Design Matrix

Questions and Subquestions	Indicators and Observable Implications	Information Sources and Data Collection Methods	Data Analysis Methods	Risks and Mitigation Strategies
EQ 1. To what extent has the 2023 World Bank Group Scorecard improved the measurement of outcomes within the institution?				
EQ 1.1. To what extent is the design of the Scorecard consistent with improving the measurement of outcomes?	Bank Group results indicators measure outcomes (not inputs or outputs) to a greater extent than comparable indicators in the previous Scorecard; indicators measure what they aim to measure. Methodologies for calculating Bank Group results indicators are replicable, aligned across Bank Group institutions, and capture all contributions adequately; data disaggregation is accurate and feasible.	• Metadata and methodology notes for Scorecard indicators • Interviews with World Bank Global Practice and IFC and MIGA focal points for Scorecard indicators	• Analysis of metadata and methodology notes to assess whether they are measuring outcomes, whether they are measuring what they set out to measure, the replicability of Bank Group results indicators, and the accuracy of data disaggregation	Because of time limitations, the replication analysis will be done only for a sample of projects for a subset of indicators.
EQ 1.2. To what extent is the Scorecard leading to improved measurement of outcomes?	Retrofitting of Bank Group results indicators of existing projects to align with the Scorecard was based on sound methodologies. Methodology for calculating forecast values for results indicators is adequate, and targets reflect a plausible trajectory of change (that is, reasonably achievable within the proposed time frame). Bank Group results indicator data sets pass basic data quality checks.	• Portfolio information on Bank Group operations for replicability • Interviews with senior management, DSOs, DE units, country directors, IFC RIDs and RIMs, GIDs, CDI, CPM, MIGA Operational Group and MIGA Economics and Sustainability Department, TTLs, investment officers, and underwriters	• Analysis of interviews to identify measurement challenges, enabling and constraining factors, and complementarities and differences across institutions • Content analysis of results narratives to assess their robustness as a complement to Bank Group indicators	Interviewees might be biased toward positive aspects of the Scorecard, making it difficult to understand challenges and opportunities for improvement. Explicit questions on the latter will be included in interview

Questions and Subquestions	Indicators and Observable Implications	Information Sources and Data Collection Methods	Data Analysis Methods	Risks and Mitigation Strategies
	Clients either have the capacity to collect Scorecard data or receive support to develop it. Compliance with Scorecard requirements does not impede other loan or advisory services and analytics preparation and supervision tasks. Results narratives complement the indicators to show integrally Bank Group contributions to development outcomes in the sectors covered by the Scorecard.	• Results narratives and interviews with their creators and users		instruments, and interviewers will be ready to pose follow-up questions. It will be important to triangulate sources.
EQ 2. To what extent do the design, implementation, and use of the 2023 Bank Group Scorecard serve its intended purposes of improving management for outcomes, accountability to the Board, and external communication about outcomes?				
EQ 2.1. To what extent are the rollout and implementation of the Scorecard managed effectively?	Scorecard implementation followed a plan, with clear objectives, roles and responsibilities, budget, and timeline. The plan is communicated to those responsible for implementing it; guidance, training, and support are provided. Feedback is absorbed, and adjustments are made as needed.	• Interview with Scorecard designers (Outcomes Department), implementers (World Bank DSOs and DE units, IFC RIDs and GIDs), and users (World Bank country and division directors, TTLs, and investment officers)	• Analysis of interviews to understand the process of rollout and implementation and understand factors enabling and constraining adequate implementation related to that process	
EQ 2.2. To what extent are the design and use of the Scorecard leading to improved management for outcomes?	Scorecard users find the number of indicators and balance between organizational effectiveness and Bank Group results indicators appropriate for management. Organizational effectiveness indicators contribute to the	• Literature on effective scorecard design; interviews with scorecard implementers and users • Prior evaluations of and research literature on Bank Group development effectiveness	• Quantitative analysis to assess the relationship between organizational effectiveness indicators and Bank Group results indicators and between Bank Group results	Interviewees might be biased toward positive aspects of the Scorecard, making it difficult to understand challenges and opportunities for improvement. Explicit

Questions and Subquestions	Indicators and Observable Implications	Information Sources and Data Collection Methods	Data Analysis Methods	Risks and Mitigation Strategies
	achievement of the Bank Group results indicators. Bank Group results indicators contribute to the achievement of client context indicators, with changes in values actionable for management purposes. Country Partnership Framework objectives, country portfolios, and project objectives reflect Scorecard priorities, including attention to groups for whom disaggregated data are now available. Budget has changed to reflect the priorities set out by the Scorecard. World Bank, IFC, and MIGA management uses the Scorecard information to set incentives (for example, in performance discussions and institutional awards). Scorecard data are used to facilitate and prioritize Bank Group joint initiatives.	• Construction of historical time series for a sample of client context and Bank Group results indicators • Budget and portfolio information	indicators and client context indicators • Portfolio review and analysis to identify changes in the composition of the portfolio after the introduction of the Scorecard • Budget analysis to assess changes in budget composition after the introduction of the Scorecard • Analysis of interviews to identify the influence of the Scorecard on project design (including components and monitoring and evaluation), changes in country engagement, and overall changes in priorities	questions on the latter will be included in interview instruments, and interviewers will be ready to pose follow-up questions. It will be important to triangulate sources.
EQ 2.3. To what extent are the design and use of the Scorecard altering the behavior of the Board or inducing it to take any corrective action?	Board members are aware of what the Scorecard measures and what it does not measure. Scorecard information is complementary to other mechanisms for Board engagement. Board members use Scorecard information in agenda-setting and strategy discussions. Board members use the Scorecard data and results narratives to	• Interviews with executive directors • Workshop and interviews with advisers to the executive directors • Records of Board discussions	• Analysis of interviews to assess familiarity with and use of the Scorecard for accountability and communication purposes • Content analysis of Board minutes to assess the extent to which the Scorecard information is used in executive directors'	

Questions and Subquestions	Indicators and Observable Implications	Information Sources and Data Collection Methods	Data Analysis Methods	Risks and Mitigation Strategies
	substantiate reports or discussions with ministries and parliaments in the countries they represent.		statements and Board discussions	
EQ 2.4. To what extent are the design and use of the Scorecard improving communication about outcomes in client and donor countries and contributing to the harmonization of results reporting across MDBs?	Scorecard dashboard format supports external communication with accessible data, simple and persuasive narrative structure and language, multimedia, and availability of data and narratives. The Scorecard allows for communication of results via secondary communication assets and avenues, for example, social media, exportable data and visualizations, provision of a mobile-friendly version, and accessibility in areas with lower internet speeds. Bank Group ECR Scorecard messaging is consistent with other messaging. Bank Group ECR facilitates the use of Scorecard narratives and outcome information by client and donor countries. Clients and donors use the narratives and Scorecard results in their own communications. Scorecard dashboards and results narratives are consulted by external audiences. Other MDBs share interest in harmonization; accept the Bank Group's convening role.	• Bank Group Scorecard website (data and visualizations, navigation, and results narratives); other MDB scorecard interfaces • Interviews with Bank Group ECR teams; review of websites and external communications of ECR and select country offices • Interviews with executive directors • Review of websites and external communications of donor and client countries • Usage statistics of the Scorecard website; information from website's feedback function • Interviews with Scorecard designers and implementers in other MDBs	• Analysis of Scorecard dashboard and website (in particular, presentation of indicators, data, visualizations, and results narratives); analysis of other MDB scorecard interfaces • Analysis of interviews; content analysis of website feedback	

Questions and Subquestions	Indicators and Observable Implications	Information Sources and Data Collection Methods	Data Analysis Methods	Risks and Mitigation Strategies
	The Bank Group Scorecard influences other MDB scorecards.			
EQ 3. Are there any risks or unintended negative consequences in the design or use of the Scorecard? What mitigation measures or complementary approaches might be put in place to address them?				
No subquestions	The Scorecard has led management to deprioritize relevant areas (in terms of attention, resources, and so on) that are not included in the Scorecard. Other unanticipated positive or negative changes in actions or behavior triggered by the Scorecard.	• Prior evaluations and research literature on scorecards and MDBs • Interviews with executive directors, management, staff, and civil society organizations • Portfolio review and analysis for a selected number of sectors	• Structured literature review focused on the unintended consequences of scorecards in MDBs and other international organizations • Analysis of interviews to identify unintended consequences • Portfolio analysis to assess the extent to which sectors included in the scorecard are prioritized and sectors not included are deprioritized	

Source: Independent Evaluation Group.
Note: CDI = Corporate Development Impact; CPM = central portfolio management; DE = Development Effectiveness; DSO = director of strategy and operations; ECR = external and corporate relations; EQ = evaluation question; GID = global industry director; IFC = International Finance Corporation; MDB = multilateral development bank; MIGA = Multilateral Investment Guarantee Agency; RID = regional industry director; RIM = regional industry manager; TTL = task team leader.

Appendix B. Sampling Frame and Sampling Strategies

Potential Informants	Sampling Frame	Sampling Frame Size	Goal	Method	Sampling Strategy
World Bank Outcomes Department	Extended leadership team and staff of the Outcomes Department	~10	Understand the thinking behind the Scorecard design and challenges found in implementation	Interviews	Census
World Bank OPCS	Staff who participated in the design of the Scorecard or who have been involved in the Scorecard's implementation	~5	Understand the thinking behind the Scorecard design and challenges found in implementation	Interviews	Census
IFC CSI—strategic initiative	Staff who participated in the design of the Scorecard or who have been involved in the Scorecard's implementation	~3	Understand the thinking behind the Scorecard design and challenges found in implementation, as well as its specific role to identify project cohort	Interviews	Census
IFC CDI—development impact	Staff who participated in the design of the Scorecard or who have been involved in the Scorecard's implementation	~7	Understand the thinking behind the Scorecard design and challenges found in implementation, as well as project cohort identification	Interviews	Census
MIGA Economics and Sustainability Group	Members of the extended leadership team and other staff who participated in the design of the Scorecard or who have been involved in the Scorecard's implementation	~10	Understand the thinking behind the Scorecard design and challenges found in implementation	Interviews	Census

Potential Informants	Sampling Frame	Sampling Frame Size	Goal	Method	Sampling Strategy
World Bank GP focal point and IFC and MIGA focal points for each of the indicators (when applicable)	Lead GP, IFC, and MIGA focal points for each of the Scorecard's indicators	~39	Understand issues related to indicator design, methodologies, and actual estimation (including retrofitting)	Interviews	Census for lead focal points and snowball sampling for other staff suggested by the lead focal points
World Bank DSOs and regional VPs	DSOs of the seven regional VPs	7	Understand challenges in the Scorecard implementation, including estimation and retrofitting of indicators, changes in project design and results frameworks, changes in country engagement, and challenges related to clients (for example, for measuring indicators)	Interviews	Census
World Bank DE units	Managers and Scorecard leads in seven regional DE units	7	Understand challenges in the Scorecard implementation, including estimation and retrofitting of indicators, changes in project design and results frameworks, changes in country engagement, and challenges related to clients (for example, for measuring indicators)	Interviews	Census
IFC regional industry directors	IFC regional industry directors and managers	12	Understand challenges in the Scorecard implementation, including data collection from clients, review of methodologies, project identification, and indicator results	Interviews	Purposive sample
IFC global industry directors	IFC global industry directors and managers	7	Understand challenges in the Scorecard implementation, including development of results narrative and review of methodologies	Interviews	Purposive sample

Potential Informants	Sampling Frame	Sampling Frame Size	Goal	Method	Sampling Strategy
Central portfolio management director	Central portfolio management director, manager, and officers	1	Understand challenges in the Scorecard implementation, including results indicator data preparation and validation	Interviews	Census
Executive directors	World Bank Group executive directors	~50 (principals and alternates)	Understand how the Board is using the Scorecard to hold Bank Group management accountable, to make policy and budgetary decisions, and to keep their constituencies informed	Interviews	Purposive sampling of executive directors with the balance between donor and recipient countries
Advisers to the executive directors	Advisers to the executive directors	~25 (assuming one with more relevant knowledge per executive director)	Understand how the Board is using the Scorecard to hold Bank Group management accountable, to make policy and budgetary decisions, and to keep their constituencies informed	Workshop and interviews	Convenience sample for workshop (that is, open invitation); purposive sample for interviews following the sampling strategy of executive directors
Division directors and country directors	All division directors and country directors	~116	Understand challenges in the Scorecard implementation, including estimation and retrofitting of indicators, changes in project design and results frameworks, changes in country engagement, and challenges related to clients (for example, for measuring indicators)	Interviews	Purposive sampling based on the following criteria: (i) countries that have been preparing CPFs since the inception of the Scorecard, (ii) IBRD and IDA countries, and (iii) countries where the engagement related to jobs will be relevant

Potential Informants	Sampling Frame	Sampling Frame Size	Goal	Method	Sampling Strategy
TTLs, investment officers, and underwriters	TTLs, investment officers, and underwriters of active projects that have been included in the estimation of the Scorecard indicators and of projects in the pipeline	~4,000	Understand challenges in the Scorecard implementation, including estimation and retrofitting of indicators, changes in project design and results frameworks, and challenges related to clients (for example, for measuring indicators)	Interviews and survey	Stratified random sample of TTLs for projects included in the Scorecard, stratified by indicator Stratified random sample of TTLs for projects in the pipeline, stratified by GP

Source: Independent Evaluation Group.
Note: CDI = Corporate Development Impact; CPF = Country Partnership Framework; CSI = Strategic Initiatives, Analytics, and Knowledge Management; DE = Development Effectiveness; DSO = director of strategy and operations; GP = Global Practice; IBRD = International Bank for Reconstruction and Development; IDA = International Development Association; IFC = International Finance Corporation; MIGA = Multilateral Investment Guarantee Agency; OPCS = Operations Policy and Country Services; TTL = task team leader; VP = vice president.

Appendix C. Structure and Evolution of the World Bank Group Scorecard

	2020 Scorecard	2023 Scorecard	Additions in the 2023 Scorecard
Development context	29 tier 1 (development context) indicators	8 vision indicators 22 client context indicators	Vision indicators: countries with high income inequality; percentage of people at high risk of climate-related hazards; percentage of people facing nutrition and food insecurity; number of hectares of key ecosystems Client context indicators: percentage of people covered by social protection and labor market programs; percentage of children who cannot read by end-primary school age; universal health coverage index; number of countries or economies at high risk of, or in, debt distress; number of countries or economies with increasing renewable natural capita per capita; percentage of terrestrial and aquatic areas that are protected; proportion of fish stocks within biologically sustainable levels; percentage of people exposed to hazardous air quality; percentage of people facing food and nutrition insecurity; percentage of people with access to reliable all-year-round transport; percentage of population using the internet; index of online e-government service provision; youth not in education, employment, or training; wage and salaried workers percent total; percentage of population in countries and economies classified as fragile and conflict-affected situations and living in extreme poverty; private investment as percentage of GDP
World Bank Group–supported results	13 tier 2 (client results) indicators for the Bank Group and 19 for the World Bank	22 Bank Group results indicators	Countries benefiting from strengthened capacity to prevent, detect, and respond to health emergencies; countries in, or at high risk of, debt distress that implemented reforms; countries with tax revenues to GDP ratios below 15 percent that have increased collections; beneficiaries with enhanced resilience to climate risks; hectares of terrestrial and aquatic areas under enhanced management; people with strengthened food and nutrition security; gigawatts of renewable energy capacity enabled; people using digitally enabled services; people benefiting from actions to advance gender equality; displaced people provided with services and livelihoods

	2020 Scorecard	2023 Scorecard	Additions in the 2023 Scorecard
Organizational effectiveness and efficiency	21 tier 3 (performance) indicators for the Bank Group and 32 for the World Bank	37 dashboard indicators (World Bank Group 2025)	Total commitments, adaptation share of climate financing, US dollars mobilized through cofinancing, countries with current Country Partnership Framework, quality of citizen and civic engagement, problem projects, crisis tool kit options, lending program delivery share, long-term finance own-account commitment, IBRD and IDA cancellations, and "scalable solutions" (share of additional financing in operations, number of multiphase programmatic approach operations, and average commitment per operation); "faster or simpler operations" (preparation time from activity initiation summary to concept review and approval, time from approval and mandate to first disbursement, and share of operations processed on track 1); "working together" (number of joint Bank Group operations); and "useful knowledge" (Country Climate and Development Reports published and number and completion rate of advisory services and analytics)
Results narratives	n.a.	15 results narratives	

Sources: Independent Evaluation Group; World Bank Group 2020, 2025.
Note: IBRD = International Bank for Reconstruction and Development; IDA = International Development Association; n.a. = not applicable.

References

World Bank Group. 2020. *Corporate Scorecards: December 2020*. World Bank Group.

World Bank Group. 2025. "World Bank Group Scorecard FY24–FY30: WBG Organizational Effectiveness and Efficiency Dashboard: Codebook—Summary and Brief Definitions." World Bank Group.

Appendix D. Structure of Scorecards and Results Frameworks in Other Multilateral Development Banks

All regional multilateral development banks have scorecards or corporate results frameworks that encompass supported outcomes and internal management indicators. Like the World Bank Group, the African Development Bank and the Inter-American Development Bank also include development context indicators (table D.1). The Asian Development Bank and the Inter-American Development Bank have updated their results frameworks since the launch of the Bank Group Scorecard.

Table D.1. Structure of Scorecards and Results Frameworks in Other Multilateral Development Banks

	AfDB (2016–25)	ADB (2025–30)	EBRD (2023)	IDB (2024–30)	World Bank Group (2023)
Development context	Total 40 Energy (5) Nutrition, agriculture (7) Private sector development (7) Regional integration (5) Quality of life (8) Cross-cutting (8)	—	—	Total 15 Poverty and inequality (5) Climate change (5) Regional growth (5)	Total 30 People (4) Prosperity (5) Planet (11) Infrastructure (2) Digital (2) Cross-cutting (6)
Results from supported operations	Total 30 Energy (8) Nutrition, agriculture (5) Private sector development (6) Regional integration (2) Quality of life (5) Cross-cutting (4)	Total 12 People (4) Planet (5) Prosperity (3)	Total 11 Transition impact (11)	Total 26 Biodiversity, climate (5) Gender, equality, inclusion (3) Institutional capacity (3) Social protection and human capital (4) Infrastructure (5) Private sector development (3) Regional integration (3)	Total 20 People (4) Prosperity (2) Planet (5) Infrastructure (3) Digital (2) Cross-cutting (4)
Operational and organizational effectiveness	Total 35 Development impact 4) Quality and speed (6) Portfolio performance (7)	Total 23 Quality of completed operations (4) Strategic alignment (10) Operational management (3)	Total 13 Operational performance (6) Financial performance (2)	Total 24 Impact orientation (4) Strategic selectivity (11) Effective instruments (3)	Total 35 Financing for outcomes (9) Monitoring for outcomes (12) Implementing for outcomes (14)

	AfDB (2016–25)	ADB (2025–30)	EBRD (2023)	IDB (2024–30)	World Bank Group (2023)
	Knowledge, advisory (1)	Organizational effectiveness (6)	Institutional performance (4)	Public private synergies (1)	
	Decentralization (2)		Resource framework (1)	Corporate foundations (5)	
	Financial performance and mobilization (4)				
	Value for money (4)				
	Staff engagement (7)				

Sources: ADB 2024; AfDB 2017; EBRD 2023; IDB 2024.
Note: ADB = Asian Development Bank; AfDB = African Development Bank; EBRD = European Bank for Reconstruction and Development; IDB = Inter-American Development Bank.

References

ADB (Asian Development Bank). 2024. *Steering ADB's Corporate Strategy to Success: Corporate Results Framework, 2025–2030*. ADB.

AfDB (African Development Bank). 2017. *The Bank Group Results Measurement Framework 2016–2025*. AfDB.

EBRD (European Bank for Reconstruction and Development). 2023. *Strategy Implementation Plan 2023–25*. EBRD. https://www.ebrd.com/home/who-we-are/strategies-governance-compliance/ebrd-strategies-policies/the-ebrd-s-strategy-implementation-plan.html.

IDB (Inter-American Development Bank). 2024. *Impact Framework 2024–2030*. IDB.

Appendix E. Outcome Definition

Outcome (compared with output) classification was discussed extensively in the Independent Evaluation Group's *Results and Performance of the World Bank Group 2020*. The Independent Evaluation Group developed a framework that is reproduced in this appendix and will be used to assess whether the outcomes described in the Scorecard are truly outcome level.

(Level 1) Outputs: Activities deliver outputs, such as knowledge products, goods, equipment, and services. From a monitoring and evaluation standpoint, outputs tend to be readily observable and countable, and their delivery can easily be tagged to a specific intervention or "attributed" to a specific organization.

(Level 2) Immediate outcomes: Outputs can have immediate effects on stakeholders or their environment. Government, private sector, and nonstate actors can gain new skills or capabilities, and citizens can have enhanced access to better-quality services or environmental benefits and see a material improvement in their living conditions. From a monitoring and evaluation standpoint, with the right approach to evidence gathering, these types of changes can be identified and often measured, and a plausible link between interventions and the observed changes can be established.

(Level 3) Intermediate outcomes: Such immediate effects can trigger further changes if they have ripple effects within a sector or if they build on the changes created by other interventions.

(Level 4) Long-term outcomes: If this is sustained, long-term outcomes eventually arise, with sustained changes in delivery, governance, or citizens' well-being.

Bibliography

World Bank. 2020a. *Results and Performance of the World Bank Group 2020*. Independent Evaluation Group. World Bank.

World Bank. 2020b. "Results and Performance of the World Bank Group 2020." Concept Note. Independent Evaluation Group. World Bank.

World Bank. 2020c. *The World Bank Group Outcome Orientation at the Country Level.* Independent Evaluation Group. World Bank.