

Project Evaluation

OVE's Review of PCRs and XSRs: The 2020 Validation Cycle

Copyright © [2020] Inter-American Development Bank. This work is licensed under a Creative Commons IGO 3.0 Attribution - NonCommercial - NoDerivatives (CC-IGO BY-NC-ND 3.0 IGO) license

(<http://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode>) and may be reproduced with attribution to the IDB and for any non-commercial purpose. No derivative work is allowed.

Any dispute related to the use of the works of the IDB that cannot be settled amicably shall be submitted to arbitration pursuant to the UNCITRAL rules. The use of the IDB's name for any purpose other than for attribution, and the use of IDB's logo shall be subject to a separate written license agreement between the IDB and the user and is not authorized as part of this CC-IGO license.

Note that link provided above includes additional terms and conditions of the license.

The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent.



© **Inter-American Development Bank, 2020**

Office of Evaluation and Oversight
1350 New York Avenue, N.W.
Washington, D.C. 20577
www.iadb.org/evaluation

RE-552
November 2020

Project Evaluation

**OVE's Review of
Project Completion
Reports (PCRs)
and Expanded
Supervision
Reports (XSRs)-
The 2020
Validation Cycle**

Office of Evaluation and Oversight



Contents

Acknowledgements	v
Acronyms and Abbreviations	v
Executive Summary	vii
Introduction	01
Project Evaluation Framework and Reviewed Projects, 2020	04
A. Public sector (SG) projects reviewed in 2020	07
B. Private sector (NSG) projects reviewed in 2020	11
Projects Reviewed by OVE, 2017-2020	18
A. Public sector operations (SG).....	19
B. Private sector operations (NSG).....	22
Lessons Learned	28
A. Review of lessons	29
B. Institutional learning	33
Conclusions and Recommendations	34
Annex I: List of SG operations reviewed	
Annex II: OVE ratings of SG operations	
Annex III: OVE PCR reviews	
Annex IV: List of NSG operations reviewed	
Annex V: NSG aggregated ratings	
Annex VI: Early operating maturity by project type	
Annex VII: Text analysis of lessons learned	

Acknowledgements

This report was led by María Fernanda Rodrigo López, under the general direction of Ivory Yong-Prötzel, OVE Director. The team included Claudia Figueroa, Oliver Peña-Habib, Xiomara Rojas-Asqui, and Melanie Putic.

Acronyms and Abbreviations

CBA	Cost-Benefit Analysis
CEA	Cost-Effectiveness Analysis
CO	Operational Closure
DEA	Development Effectiveness Analytics
DEF	Development Effectiveness Framework
DELTA	Development Effectiveness Learning, Tracking, and Assessment Tool
DEM	Development Effectiveness Matrix
DEO	Development Effectiveness Overview
DIAS	Development Impact Assessment Scorecard
EOM	Early Operating Maturity (NSG projects)
FI	Financial Intermediary
IDB(G)	Inter-American Development Bank (Group)
ICC	Inter-American Investment Corporation (since 2017 IDB Invest)
M&E	Monitoring and Evaluation
NPL	Non-Performing Loans
NSG	Non-Sovereign-Guaranteed
OMJ	Opportunities for the Majority

OVE	Office of Evaluation and Oversight
PBL	Policy-Based Loan
PBP	Policy-Based Programmatic Loan
PCR	Project Completion Report
PMR	Project Monitoring Report
SCF	Structured and Corporate Finance Department
SG	Sovereign-Guaranteed
SMEs	Small and Medium-Sized Enterprises
XSR	Expanded Supervision Report

Executive Summary

This report summarizes the results of the Office of Evaluation and Oversight's (OVE) annual validation of the self-assessments of project performance and results completed by the Inter-American Development Bank (IDB) and IDB Invest in 2019–2020. The IDB and IDB Invest have systems in place to measure the development effectiveness of their operations. These systems use a number of instruments to assess projects at the design, implementation, and completion phases. The design-phase assessment uses a “Development Effectiveness Matrix” (DEM) for sovereign-guaranteed (SG) operations. The effectiveness of non-sovereign-guaranteed (NSG) operations is measured with a “Development Effectiveness Learning, Tracking, and Assessment tool” (DELTA). Implementation-phase assessments measure SG operations with “Project Monitoring Reports,” while NSG operations are assessed with “Project Supervision Reports.” Upon completion, SG projects are self-evaluated with “Project Completion Reports,” or PCRs. NSG operations, use “Expanded Supervision Reports,” or XSRs when they reach early operation maturity (EOM). OVE validates Management’s self-evaluations (PCRs and XSRs) and assigns a final project performance rating to each operation.

As part of the 2020 validation cycle, OVE reviewed PCRs for 63 operations, 62 with operational closure (CO fully justified) in 2018 and one in 2013. XSRs were reviewed for 36 IDB Invest operations that had reached early operating maturity (EOM) in 2018. The PCRs for the 63 operations validated covered 54 investment operations, five policy-based programs (PBPs), three policy-based Loans (PBLs), and one reimbursable technical cooperation spanning 21 IDB member countries and representing a total approved volume of USD\$5.6 billion. The 36 XSRs covered 22 NSG operations that had been approved by the former IIC (now IDB Invest), five by IDB’s Opportunity for the Majority Initiative, and nine by IDB’s Structured and Corporate Finance Department. Together they represent an approved total volume of USD\$566 million and cover operations in 15 IDB Invest member countries and one regional operation across three countries. Among the NSG operations, financial intermediary (FI) operations dominated with 28 of the 36 operations, totaling USD\$537 million (95%) of approved loan volumes.

The IDB Group uses an objectives-based methodology to evaluate the performance of its lending operations. This methodology measures project performance against the development objectives stated at project approval. It assesses the relevance of these objectives, and the associated project designs; to what extent the objectives were achieved; how efficiently project resources were used; and how sustainable the results are. The four core criteria (relevance, effectiveness, efficiency, and sustainability) are examined and then rated on a four-point scale ranging from excellent to unsatisfactory. Based on these core criteria, each project is rated for overall outcome, which is a weighted average of the four core criteria: relevance, efficiency, and sustainability criteria are each weighted 20%, while effectiveness is 40%. The overall outcome rating of the project is rated on a six-point scale, from highly successful to highly unsuccessful. Besides the core criteria, the 2018 guidelines require the inclusion of non-core criteria. Although these criteria are rated, they are not included in the project's overall outcome rating. PCRs assess the performance of the Bank and its counterparts (i.e., Borrowers). The XSRs include the financial and nonfinancial additionality of IDB Invest, the investment outcome, and IDB Invest work quality.

For simplicity, this report groups ratings into a binominal summary rating (i.e., positive and negative). The term 'positive' refers to overall outcome ratings judged 'partly successful,' 'successful,' and 'highly successful,' as well as core criteria ratings of 'satisfactory' and 'excellent.' The term 'negative' refers to overall outcome ratings of 'partly unsuccessful,' 'unsuccessful,' and 'highly unsuccessful,' and core criteria ratings of 'partly unsatisfactory' and 'unsatisfactory.'

In the 2020 validation cycle, overall project outcome ratings were positive for 51% (32) of SG operations and 58% (21) for NSG operations. In terms of core criteria, as in previous cycles, the highest ratings were for relevance (84% positive for SG and 61% for NSG), showing strong vertical logic and alignment with the IDBG strategies and country realities. Effectiveness had the lowest ratings (29% positive for SG and 50% for NSG), which drove overall development outcomes down. For SG operations, a combination of factors explains negative effectiveness ratings such as poor monitoring and evaluation (M&E), cancellation of outputs or components that prevented outcomes from materializing, or underperformance. Negative ratings on effectiveness for NSG operations stemmed mostly from unmet objectives. XSRs frequently hypothesize that failures to reach objectives were related to factors outside the control of the client (e.g., economic crises, changes in regulatory environment, decreased demand

for SME credits), but they rarely provide a clear analysis of how project design or supervision fell short on taking such external factors into account.

This report also expands on overall project outcome ratings and core criteria for all validation completed between 2017 and 2020, along with an analysis of documented lessons learned in the PCRs and XSRs. To date, across these four validation cycles, OVE has validated more than 300 projects. This analysis seeks to shed light on the reasons behind the systematic low effectiveness ratings, and summarizes lessons learned from all PCR's and XSRs that could be incorporated into future operations to foster institutional learning.

To date, OVE has reviewed 183 PCRs prepared for operations with a DEM in accordance with the unified objectives-based methodology. A systematic analysis based on all validation findings suggests that M&E is a key factor affecting project ratings on overall outcome. Fifty-eight percent (106) of validated SG operations obtained a positive overall outcome rating. Although most negative overall outcome ratings involved at least two core criteria rated negative, moderately successful overall outcome results are mainly driven by low ratings on effectiveness. Negative effectiveness ratings arose from a combination of factors ranging from poor M&E quality to cancellations of outputs and underperformance of projects.

Cancellation of components or outputs for SG operations also explains negative ratings on effectiveness. The systematic analysis from all validations reveals that partial or complete cancellation of components or outputs altered the vertical logic of projects, leading to negative effectiveness ratings. This type of cancellations materialized in almost a third of projects rated negative on effectiveness (37 out of 124 cases). This raises the question why projects whose original vertical logic was altered and which thus could no longer achieve their intended objectives were continued rather than formally restructured.

OVE's validation of project outcome ratings contributes to ensure the credibility of the self-evaluation system. This is one of its key purposes. For comparable PCRs, in 2019, 46% (25 out of 54 cases) of Management's self-evaluation ratings differed from OVE's overall project outcome ratings; in 2020 it was 48% (30 out of 62 cases). Most differences resulted in a lower overall project outcome rating from OVE. Overall, while OVE found that 56% of SG operations achieved an overall positive outcome rating, Management assigned positive ratings to 75% of operations. For NSG operations, the number of cases rated differently by Management and OVE decreased from 60% in 2018 to 39% in 2020. The cases with positive ratings from Management but negative re-ratings from OVE also decreased during the period,

from 21% (9 out of 43 cases) in 2018 to 6% (2 out of 36 cases) in 2020. For both SG and NSG validations, effectiveness was the criterion adjusted most frequently by OVE, followed by efficiency.

Progress in the quality of self-evaluations notwithstanding, there is still room for improvement. The relevance criterion of PCRs requires assessments not only of a project's alignment with country development needs and Bank strategies, but also with realities in the countries for which these projects were designed (i.e., economic, environmental, social, political economy, and capacity conditions). Although relevant for an operation's vertical logic, its implementation and risk analysis, this information is often meager in PCRs. Furthermore, as in past validation cycles, information on environmental and social safeguards performance continues to be scarce. For XSRs, underachievement of development objectives has been a central issue, particularly for FIs and, when reasons are provided, they tend to refer to external factors.

Neither PCRs nor XSRs focus enough on institutional learning. The usefulness of these self-evaluation reports as learning tools, and indeed of their validations, is hampered by their somewhat limited analysis of factors affecting projects' performance. In addition, they tend not to provide clear and consistent narratives across evaluative criteria—narratives that would help to identify the drivers of a project's results. For example, XSRs relevance and work-quality challenges arose from ill-defined goals and other design flaws; while low effectiveness ratings mostly stemmed from unmet targets, rather than inadequate targets or other results matrix problems. The XSRs rarely touch on these issues.

Lessons drawn from PCRs and XSRs reveal the need to (1) strengthen IDBG's M&E practices and data collection and (2) document concrete successful actions (for projects to replicate) and unsuccessful actions (for projects to avoid). Most lessons focus on how best practices in M&E should have been, but were not, implemented, and how multiple problems with data collection hindered the calculation of indicators. But the parts that focus on lessons from implementation, aimed at producing new knowledge, are limited. This reveals the need for operations teams and the clients to adopt better M&E practices and data-collection mechanisms. The lessons learned also expose the need to document concrete actions, both those that proved ineffective (to be avoided in the future) and those that proved effective (to be replicated).

Considering these conclusions, OVE recommends:

A. For IDB Management

- 1. Ensure that all future PCRs submitted to OVE integrate alignment with country realities in the relevance assessment.** Most projects with negative relevance ratings were affected by weak vertical logic, which in turn was due to a poor diagnostic assessment or poor alignment with country realities. This then often resulted in implementation problems or cancellation of components, ultimately affecting project results. Therefore, to strengthen the learning component of the PCRs, assessments must consider country realities.
- 2. Ensure that the 2020 PCR guidelines ask for lessons relevant for institutional learning, focused on project elements to replicate or avoid.** To promote institutional learning from SG operations, lessons learned sections of PCRs should include a reflection on successful implementations (e.g., actions that built resilience under unexpected circumstances) and unsuccessful ones. Many lessons stressed the risks of implementation and recommended a more rigorous ex-ante project design to prevent these in future operations. Although a more rigorous project design is undoubtedly necessary, unexpected occurrences are often the rule, not the exception, especially in unstable political and business environments. We seem to know more about how projects should have been designed, but less about what projects did to avoid disappointing outcomes when contexts changed.

B. For IDB Invest Management

- 3. Strengthen the learning component of XSRs, incorporating a reflection on what worked and what went wrong.** According to the XSRs guidelines, the reports are a tool for accountability and learning. However, their learning aspect needs to be reinforced. While low effectiveness ratings reflect unmet objectives, XSRs have focused on analyzing indicators without valuing the need to understand what affects them and how IDB Invest might have responded. The same applies for the lessons-learned section, which mostly focuses on applying already known M&E practices rather than generating knowledge on what worked and what did not.
- 4. Adjust XSR guidelines to ensure that achievements are measured by progress against targets relative to their baseline.** Current guidelines state achievement thresholds based on the absolute proportion of the target

not considering baselines. This could lead to irrational results such as an achieved result being rated positive even if it is below the baseline. OVE recommends that the guidelines be adjusted to reflect the practice that has been initiated during the recently started 2020/2021 XSR exercise which measures progress against targets relative to their baseline.



01

Introduction

1.1 This report summarizes the results of OVE’s annual review of the self-assessments of project performance and results completed by the Inter-American Development Bank (IDB) and IDB Invest. Both IDB and IDB Invest have systems in place to measure the development effectiveness of their operations. At the IDB, the Development Effectiveness Framework (DEF), launched in 2008, consists of three assessment tools for projects from design through completion:

- The Development Effectiveness Matrix (DEM) assesses, at design stage, a project’s evaluability, or ability to report on results.
- The Project Monitoring Report (PMR) follows activities and outputs throughout the execution of the project to ensure they are produced within the expected costs and timeframes.
- The Project Completion Report (PCR) presents a self-evaluation at completion.

The DEF applies to sovereign-guaranteed (SG) operations and, until the IDBG’s “merge-out” in 2016, to non-sovereign-guaranteed (NSG) operations approved by the former Structured and Corporate Finance Department (SCF) and the former Opportunities for the Majority initiative (OMJ). On the other hand, the former Inter-American Investment Corporation (IIC) tracked development results since 2001 (CII/RE-1) by self-evaluating its operations following the Good Practice Standards for private sector operations issued by the Multilateral Development Banks’ Evaluation Cooperation Group. After 2016, IDB Invest implemented the Development Effectiveness Learning, Tracking, and Assessment tool (DELTA) for NSG operations, which has three tools that resemble the DEF instruments: (1) an ex-ante assessment of the project’s evaluability (DELTA),¹ (2) monitoring reports on project implementation, and (3) a final self-evaluation report (the Expanded Supervision Report, XSR) on the results of operations once they reach early operating maturity (EOM).²

1.2 PCRs and XSRs are self-evaluations of development operations and are intended to serve as accountability and learning tools. According to both PCR and XSR guidelines, “the accountability goal addresses the need for the IDBG to ensure that the project resources were used for the objectives for which the project was granted... The learning goal aims to replicate successes and avoid mistakes in the

1 The DELTA also replaced the evaluation system of the former Inter-American Investment Corporation (IIC) established in 1999 (CII/GN-141, CII/GN-141-2).

2 The definition of early operating maturity varies by project type. See Annex VI.

future by providing lessons to guide the execution of ongoing projects and the design of future ones” (OP-1696-5; 2018 XSR Guidelines for IDBG Private Sector Projects).

- 1.3 The IDBG’s Office of Evaluation and Oversight (OVE) supports accountability and learning through the validation of completed PCRs and XSRs. OVE contributes to the credibility of the IDBG’s project performance reporting system through the validation of PCRs and XSRs. When Management submits PCRs and XSRs, it assigns performance ratings to projects, based them on several criteria. Then OVE reviews the project’s performance information and validates these ratings, sometimes agreeing with Management’s ratings and sometimes assigning lower or higher ratings based on OVE’s assessment of project performance based on evidence presented in the self-evaluations. OVE’s project performance ratings are final and are reported in the Development Effectiveness Overview (DEO). The year 2020 is the fourth consecutive year for OVE validation of these self-evaluations in the context of the IDBG’s objectives-based evaluation framework for public and private sector operations.
- 1.4 This report presents the results of the 2020 review and validation process for PCRs and XSRs completed in 2019 and 2020. During the 2020 review exercise, OVE reviewed and validated PCRs for 63 operations that were financed by 78 IDB loans and grants. 62 operations had operational closure (CO fully justified) in 2018 and one in 2013; 36 XSRs were submitted to OVE in 2020 for 38 A loans and 1 equity investment that had reached EOM in 2018. Some PCRs and XSRs assessed a group of related projects rather than on a single individual project.
- 1.5 In addition, this report presents a summary of overall results from OVE’s last four validation cycles (2017–2020). To date, after 4 validation cycles in accordance with the unified objectives-based methodology, OVE has reviewed and validated the self-evaluations for 183 SG operations³ and 144 NSG operations. This report explores the validation results from all four years

3 The number of PCRs and XSRs is not equivalent to the number of lending operations, as a self-evaluation can cover multiple lending operations. On the SG side, a separate PCR is prepared for each individual Investment Loans, Policy-Based Loans (PBLs), stand-alone Reimbursable Technical Cooperation (RTC), and stand-alone Investment Grants (IGR). Programmatic Policy-Based Loans (PBP), Multi-Phase Programs, and Conditional Credit Lines (CCLIP) that finance dependent operations in the same sector, are covered in a single PCR that evaluates the series as a whole ([OP-1696-5](#)) and are counted as a single operation for purposes of this report. In the case of a horizontal CCLIP with independent operations, such as the PROFISCO program in Brazil, the PCR covered seven independent operations (BR-L1165, BR-L1202, BR-L1207, BR-L1235, BR-L1236, BR-L1238, and BR-L1239) and each one was validated and counted by OVE as a separate operation. On the NSG side, a series of investments with the same client and similar objectives, or regional operations are generally evaluated in a single XSR. If multiple operations covered in the same XSR have separate results matrices, OVE normally validates and counts them individually.

to complement the 2020 results, to expand on the projects' performance drivers from a longer-term perspective that may provide useful insights for future operations.

- 1.6 Finally, to foster institutional learning, this report also analyzes the knowledge generated by the “lessons learned” sections in PCRs and XSRs. Each PCR and XSR includes a section of findings and recommendations and lessons learned respectively. Institutional learning is not only an objective of the reports. It is key for development effectiveness. Reporting on the lessons learned can play an important role in improving outcomes. Staff should share successes (for replication) as well as mistakes (to avoid). OVE analyzed the lessons learned from all validated PCRs and XSRs from 2017 to 2020 to identify the main and most frequent messages.
- 1.7 This report includes five sections. Following this introduction, Section II presents the results of OVE’s validations for SG and NSG operations in 2020. Section III contains the review and overall results of all validation cycles, while Section IV analyzes lessons learned. Finally, Section V presents conclusions and recommendations.



02

Project
Evaluation
Framework
and Reviewed
Projects, 2020

2.1 The IDB Group employs an objective-based methodology for evaluating project performance. The evaluation follows the Project Completion Report (PCR): Principles and Guidelines (document [OP-1696-5](#)) for public sector operations, and, for private ones, the XSR Guidelines for IDBG Private Sector Operations, both revised in 2018. Project performance is measured against the development objectives as stated at project approval and reflected in the results matrix, assessing how relevant these objectives and the associated project designs were, to what extent the objectives were achieved, how efficiently project resources were used, and how sustainable the achieved results are. Four core criteria are examined: relevance, effectiveness, efficiency, and sustainability, each rated on a four-point scale ranging from excellent to unsatisfactory. Based on these core criteria, each project also receives an overall outcome rating calculated as a weighted average, where relevance, efficiency, and sustainability criteria weigh 20% each and effectiveness 40%. The overall rating of project outcome uses a six-point scale, from ‘highly successful’ to ‘highly unsuccessful.’ Besides the core criteria, the 2018 PCR and XSR guidelines require the inclusion of non-core criteria, which are rated but not included in calculating a project’s overall outcome rating. While PCRs include the performance of the Bank and its counterparts (i.e., Borrowers), XSRs include the financial and nonfinancial additionality of IDB Invest, the outcome of the investment, and IDB Invest work quality. (Table 2.1 presents each core criteria dimension and Table 2.2 presents non-core criteria.)

Table 2.1. Core criteria for project performance assessment

SG operations	Core Criteria	NSG operations
<ul style="list-style-type: none"> • Alignment of project development objectives with country development needs. • Alignment with IDBG country strategy. • Alignment of project design with country realities. • Alignment of project design with project development objectives. 	Relevance	<ul style="list-style-type: none"> • Alignment of project objectives with country development needs. • Alignment with IDBG country strategy and corporate goals. • Alignment of project design with country realities. • Alignment of project design with project development objectives.
<ul style="list-style-type: none"> • Extent to which project achieved each stated development objective, given project outputs produced. 	Effectiveness	<ul style="list-style-type: none"> • Extent to which project achieved each stated development objective, given project outputs produced.
<ul style="list-style-type: none"> • Extent to which project benefits exceed project costs or extent to which project benefits were achieved at less than expected or at reasonable costs. 	Efficiency	<ul style="list-style-type: none"> • Financial performance: Project contribution to company financial results and extent to which project process and business objectives were achieved. • Economic performance: Extent to which project’s economic benefits exceed costs of capital; project effects on key economic stakeholders.
<ul style="list-style-type: none"> • Safeguards performance. • Assessment of risks to continuation of project development results. 	Sustainability	<ul style="list-style-type: none"> • Safeguards performance. • Assessment of risks to continuation of project results.

Project Development Outcome

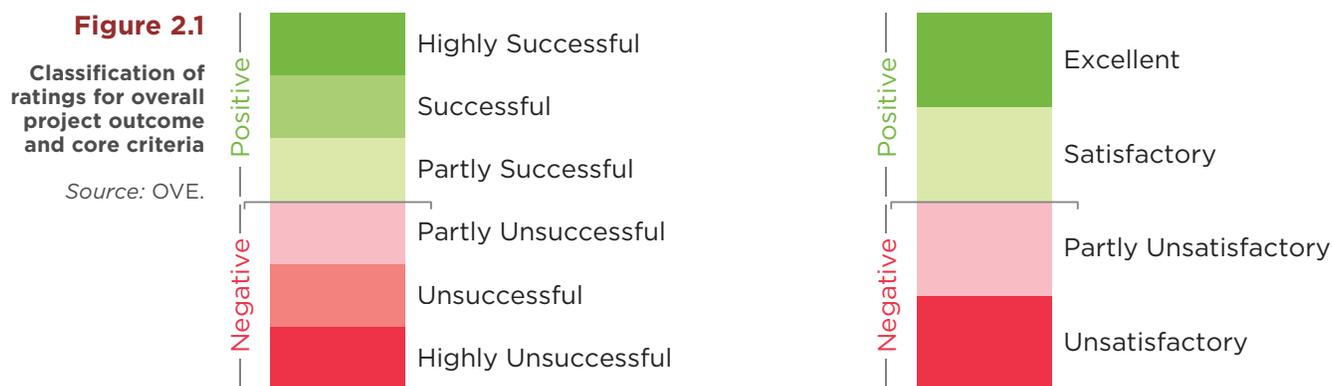
Source: OVE’s Review of Project Completion Reports (PCRs) and Expanded Supervision Reports (XSRs)—2018–2019 Validation Cycle, p. 3.

Table 2.2. Non-core criteria for project performance assessment

SG operations	Non-Core Criteria	NSG operations
<ul style="list-style-type: none"> Bank performance in ensuring quality at entry: the extent to which the Bank identified, facilitated preparation of, and approved the operation such that it was most likely to achieve planned development outcome. Quality of Bank supervision: extent to which the Bank proactively identified and resolved threats to the achievement of relevant development outcomes. 	<p>Bank performance (SG) / IDB Invest work quality (NSG)</p>	<ul style="list-style-type: none"> IDBG's work quality, taking into consideration precommitment work at entry, i.e. screening, appraisal, and structuring/underwriting. IDBG's work quality, taking into consideration its monitoring and supervision of the operation following commitment or guarantee issuance.
<ul style="list-style-type: none"> Effectiveness by which the Borrower discharged its responsibilities' including: i. compliance with covenants, agreements, and safeguards; ii. provision of timely counterpart funding; iii. measures to establish the project's sustainability, and iv. compliance with the monitoring and evaluation plan, among others that the PCR team deems relevant. 	<p>Borrower performance (only SG)</p>	
	<p>Additionality (only NSG)</p>	<ul style="list-style-type: none"> The assessment and rating is based on the counterfactual assessment of how the project would have (or would not have) proceeded without IDBG's support. It considers financial and nonfinancial additionality.
	<p>Investment outcome (only NSG)</p>	<ul style="list-style-type: none"> This section assesses the investment's gross profit contribution (net of financing costs and loss provisions but before deducting administrative costs) with rating benchmarks set in relation to corresponding at-approval standards for minimally expected performance.

Source: OVE.

2.2 For simplicity, this report groups ratings into a binominal summary (i.e., positive and negative). The term 'positive' refers to overall outcome ratings of partly successful, successful, and highly successful, as well as core criteria ratings of satisfactory and excellent. 'Negative' ratings are conferred by overall outcome ratings of partly unsuccessful, unsuccessful, and highly unsuccessful, in addition to core criteria ratings of partly unsatisfactory and unsatisfactory. (Figure 2.1 shows the classification of ratings for overall project outcome and core criteria ratings.)



A. Public sector (SG) projects reviewed in 2020

2.3 OVE reviewed PCRs for 63 operations in 2020 of which 62 closed in 2018 and one in 2013.⁴ Only one PCR for a loan with operational closure 2018 was not submitted on time.⁵ The PCRs for the 63 operations validated covered 54 investment operations, five policy-based programs (PBPs), three policy-based loans (PBLs), and one reimbursable technical cooperation. The total volume of approved loans and grants covered by the validation cycle was USD\$5.63 billion. Table 2.3 presents the breakdown of validated PCRs by department and the volume of loans/grants approved. (Annex I presents all validations).

Table 2.3. Distribution of validated PCRs by sector

Department	SG operations* (no. of PCRs)	Total approvals** (USD\$ billions)
Institutions for Development	19	2.48
Infrastructure and Environment	16	1.62
Social	15	1.11
Climate Change & Sustainable Development	11	0.41
Integration and Trade	2	0.1
Total	63	5.63

*Five of these operations correspond to PBP series that cover 13 programmatic loans; three correspond to investment loans that were accompanied by an investment grant each, all from water and sanitation; one corresponds to a multiphase program of three investment loans; one is a reform loan with a refundable technical cooperation; and one an investment loan with supplementary resources.

**Includes the approved amount of the entire operation.

Source: OVE.

1. Overall project outcomes

2.4 The overall development outcome rating summarizes the overall performance of a project derived from the core criteria of relevance, effectiveness, efficiency, and sustainability). Of the 63 validated SG operations, 51% (32) obtained a positive OVE rating for overall project outcome. Two of these were highly successful and corresponded to investment projects in the social sector (education and health). Nearly half the validations

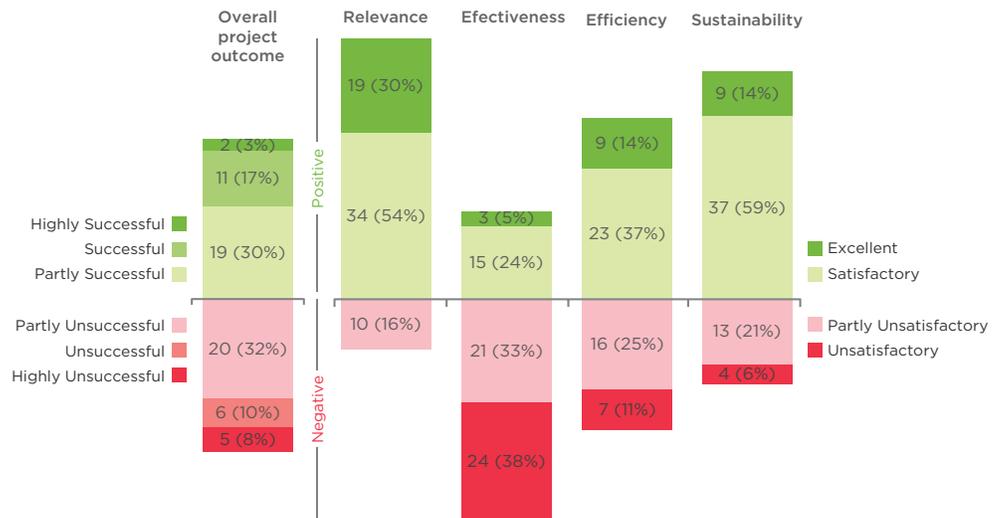
4 73% of these operations were approved before 2013 with 2009, 2010, and 2012 accounting for 20% of approvals each.

5 OVE's project performance ratings are final and are reported in the DEO (Development Effectiveness Overview). OVE must submit final project performance ratings to Management before the DEO's is sent to the President's committee (PCY) for approval. This year, the deadline for OVE to submit final ratings was May 24. The PCR for project CO-L1019, San Francisco–Mocoa Alternate Road Construction Project—Phase I, was approved only in late August 2020. Neither this report nor the DEO 2020 includes its results.

(31, or 49%) resulted in a negative rating for overall project outcome. Although all these projects had negative ratings for effectiveness, more than three-quarters of them had at least one other negatively rated criterion. The most frequent combination for negative ratings was effectiveness and efficiency, followed by effectiveness and sustainability. Furthermore, 6 of these negatively rated projects were marked by cancellations affecting at least 40% of their original approved resources. Out of the 31 negatively rated operations, 5 were ‘highly unsuccessful’ and 6 ‘unsuccessful.’ ‘Highly unsuccessful’ operations featured design flaws (e.g., lack of alignment with country realities, weak vertical logic, overly complex implementation schemes) that affected relevance, or limitations in the M&E plans (e.g., lack of information on indicators at project start or end, or challenges in attribution of results) that lowered the effectiveness rating. Project implementation was also impeded by counterparts’ performance when the latter did not fully comply with institutional arrangements. For example, the executing units had insufficient technical capacity or personnel for project implementation or institutional coordination was weak.

Figure 2.2
Number and share of SG project ratings: Overall outcome and core criteria

Source: OVE.



2. Core criteria: Relevance, effectiveness, efficiency, and sustainability

2.5 Among the core criteria, projects reviewed in 2020 scored highest on the relevance ratings in 2020. As in previous validation cycles, projects scored highest on the relevance criterion. In 2020, 84% (53) of validated operations were rated positive on relevance. In other words, most operations were aligned with country’s development challenges, the Bank’s strategies, the country’s realities, and had adequate vertical logic. The less relevant operations, despite being aligned with the Bank’s strategies, had weak vertical logic which in 7 out of 10 cases was affected by inadequate alignment with country

realities (i.e., the objectives and design of the intervention did not consider the economic, environmental, social, political economy, or capacity conditions in place) (e.g., AR-L1078, BH-L1030, BR-L1084, ES-L1050).⁶

- 2.6 The effectiveness criterion had the lowest ratings. In 2020, only 18 operations (29%) were rated positive on effectiveness. On average, these operations had achieved over half of their development objectives (i.e., they met at least 80% of desired outcomes associated with each of the objectives achieved relative to their baseline, and no objective was rated unsatisfactory). Conversely, 45 operations (71%) had negative ratings for effectiveness. Low effectiveness ratings are explained by a combination of factors, the most frequent ones are related to (i) the quality of M&E (30 cases);⁷ (ii) poor performance (of 18 cases, 9 were explained exclusively by underperformance); and (iii) cancellation of outputs that prevented achievement of outcomes (of 16 cases, seven reported cancellations of more than half their original approved resources).⁸
- 2.7 More than half of the validated investment projects proved cost-effective/ efficient. Fifty-eight percent (32 of 55)⁹ of the validated projects obtained positive scores for efficiency.¹⁰ In these cases, the analyses were robust and incorporated at least 50% of the total program costs. Of the 23 projects with negative efficiency ratings, nine failed to present cost-benefit or cost-efficiency analyses,¹¹ five presented inadequate assumptions,¹² and six performed partial analyses (i.e., incorporated less than 50% of program costs). The remaining three projects turned out not to be cost-effective/efficient. Out of the 12 cost-effectiveness analyses presented in the PCRs, eight did not clearly document

6 At the design phase, a strategic mapping of the relevant stakeholders was not carried out and negotiation strategies were not developed with all of them before implementation; neither activities were foreseen to strengthen the regulatory framework in order to reduce the risks associated with the lack of an agreed implementation plan with relevant stakeholders.

7 Related to the quality of the results matrix (i.e., indicators are not SMART, misuse of outputs as results indicators, lack of definition of indicators, baselines, or goals); and lack of data at end of project (i.e., availability issues with the data related to the context of the project, noncompliance with the M&E plan caused by budget shortfalls or faulty implementation of data collection, or the executing unit's capacity to collect data).

8 For example, for the project BH-L1030, there is no data on the outcome indicators for two of its three objectives (the component of conditional transfers was not implemented, so outputs were not delivered).

9 The efficiency analysis is not performed for PBLs and PBPs.

10 A positive rating for efficiency is achieved by a project when its economic rate of return is at least 80% of the discount rate used in the ex-post evaluation or it proves to be more cost-efficient than other projects with similar solutions.

11 Instead, they performed a cost- and time-overrun analysis (CTOA), with this methodology, the maximum possible score is 'partly unsatisfactory.'

12 For example, CBAs that presented benefits that could not be attributed to the program, or CEAs that used comparators that lacked external validity.

the assumptions of the analysis, such as the link between costs and outputs considered, or a discussion of the external validity of comparators used.

2.8 Validations showed that the results of 46 operations (73%) are likely to be sustainable. Almost three-quarters of the operations (46) showed the likely continuation of outcomes in which the risks to continuity were largely mitigated. The remaining operations (27%, 17) have operational costs that present budgetary challenges; in some cases, there are challenges to institutional capacities for the maintenance or operation of outputs (AR-L1045, HA-L1035). Discontinuity of the operation owing to change of government was a risk in few cases (CH-L1034, GY-G1003, HA-L1035). On the environmental and social impacts, the PCRs generally lacked information on compliance with the Bank's safeguards policy beyond the mention that guidelines of this policy were followed.¹³

3. Non-core criteria: Bank and borrower performances

2.9 The Bank's performance was positive in 45 operations (71%). Cases of low performance ratings (18 out of 63) are mostly explained by low-quality operations at the design stage, with overly ambitious implementation arrangements or designs poorly aligned with country realities, poor risk analysis, or deficient M&E.

2.10 Borrowers' performance was positive in 41 operations (65%). The challenges of responding to operations' needs at implementation lowered scores for this criterion. Ensuring adequate institutional arrangements stand out as the most frequent challenge—that is, the need to maintain endowed, stable executing units with relevant technical capacities, in addition to leadership on interinstitutional coordination. Uneven compliance with, for example, M&E responsibilities and project ownership explained some of the negative results (7 cases) in this criterion.

4. PCR quality

2.11 About half the PCRs had a positive (good or excellent) rating on quality, indicating that they analyzed the project achievements well, they were candid and complete, and ratings were consistent with evidence. Conversely, 51% of PCRs achieved a fair or poor rating. Areas for improvement for PCRs are detailed in Box 2.1. As in the previous validation cycle, several PCRs contained differences between the data reported in the PCR document and the PCR checklist, or mistakes in the calculation of overall project outcomes in the PCR checklist. The discrepancies

13 For further reference refer to document [RE-544](#), CII/RE-58 ¶3.12-3.13 p.9-10 and ¶5.7 p. 24

point to the continued need to ensure that these documents are harmonized and that Management ratings are clearly and systematically derived from content in the main text.

Box 2.1. Areas for improvement for PCRs

Criterion	Area of improvement
Relevance	<p>i. A detailed discussion of country realities: Describe how well the objectives and design of the project considered the economic, environmental, social, political economy, and capacity conditions in which it took place.</p> <p>ii. A detailed discussion of relevant changes to the vertical logic of the project during implementation, if it applies (i.e., due to canceled components or outputs).</p>
Effectiveness	i. Harmonization of data between the PCR document and the PCR checklist.
Efficiency	i. A detailed, comprehensive presentation of the efficiency analysis, in particular for CEAs.
Sustainability	i. A comprehensive discussion on safeguards performance providing means of verification for implemented measures.
Lessons learned	Include lessons that foster institutional learning (see Section IV).

B. Private sector (NSG) projects reviewed in 2020

2.12 OVE reviewed 36 XSRs that corresponded to 38 A loans and 1 equity investment, totaling an approved amount of USD\$566 million. The validations in this cycle correspond to operations that reached early operational maturity (EOM)¹⁴ in 2018. Most operations (22, or 61%) belong to the former Inter-American Investment Corporation (IIC), followed by the Department of Structured and Corporate Financing (SCF) of the IDB (9, or 25%) and by the Opportunities for the Majority initiative (OMJ) of the IDB (5, or 14%).¹⁵ The operations evaluated were approved between 2011 and 2015 with a higher concentration in the last year.¹⁶ Operations with financial intermediaries (FIs) represent most of the portfolio (78% of the number of operations).

14 The exact definition of EOM varies by project type, but implies that IDBG has made its financial material disbursement and received at least one set of audited financial statements covering 12–36 months of operating revenues post disbursement/project completion, with the period depending on the type of project. See Annex VI.

15 In other words, all operations assessed correspond to legacy projects. Operations approved by IDB Invest are not included, since they have not yet reached EOM.

16 Thirty-three percent of the resources and 47 percent of the operations were approved in 2015. Specifically, those of OMJ and SCF were approved between 2013 and 2015 and, therefore, all had a Development Effectiveness Matrix (DEM) and, in some cases, a results matrix. IIC's operations evaluated in this cycle were approved between 2011 and 2015 and included the development impact assessment scorecard (DIAS) that assigns a development impact and additionality score.

Table 2.4. Distribution of XSRs by segment and window

Sector	Private sector windows (# of XSRs)			Total	Total approvals (US\$ Millions)
	Former CII	OMJ	SCF		
Financial Intermediaries	14*	5	9	28	537.6
Corporates	6	-	-	6	24.1
Infrastructure and Energy	2	-	-	2	4.6
Total	22	5	9	36	566.2

Source: OVE, with data from Maestro and Analitika's "Maestro All Transactions" Report. Total differs by a decimal due to rounding.

1. Overall project outcome

2.13 Fifty-eight percent of NSG operations had positive ratings for overall project outcome. In total, there were 15 NSG operations with negative overall project outcome ratings, all of them exhibited deficiencies in at least two core criteria (efficiency and either effectiveness, or sustainability); 12 (80%) had problems across these three criteria. Among FI operations, the largest segment of reviewed operations, 17 of 28 operations (61%) had a positive overall project outcome rating.

Figure 2.3

Share of NSG project ratings: Overall project outcome and core criteria

Source: OVE.



2. Core criteria: Relevance, effectiveness, efficiency, and sustainability

2.14 A bit less than two-thirds of operations (or 61%) were relevant. Negative relevance ratings stemmed from design challenges. Twenty-two operations were rated satisfactory and 14 partly unsatisfactory on relevance; no operations were rated excellent or unsatisfactory. Negative ratings for relevance were related mostly to flaws in project design, frequently because of poorly defined goals (9 of the 14 cases with negative relevance rating),

with inadequate targets (underestimated, not specific to the target portfolio or to the country reality). Other less frequent reasons behind the low relevance ratings were marginal alignment with development needs¹⁷ or country realities, and choice of client, size of the loan, and insufficient diagnostics.

2.15 Half the operations achieved positive ratings for effectiveness. Negative ratings reflect insufficient achievement of project objectives, lack of attribution of project outcomes, or insufficient information to verify progress.¹⁸ In most cases, negative ratings were due to insufficient achievement of goals, meaning that on average, these projects did not achieve the majority of their objectives (i.e., they met less than 80% of desired outcomes associated with each of the objectives, or one objective was rated unsatisfactory). For more than half these operations, the evidence shows that the operation's situation at EOM worsened compared to the baseline.¹⁹ XSRs frequently ascribe failures to meet targets to factors outside the control of the client, such as unfavorable macroeconomic conditions. However, this explanation is often presented more as hypothesis and little emphasis is placed on lessons to manage or ways to adjust to external factors. In three cases, negative ratings are explained by lack of information, which hindered assessments on progress—either some objectives lacked indicators in the results matrix from approval, the baseline was not defined, or the information was not collected at the project's conclusion. Three other cases presented both effectiveness issues—unmet targets and insufficient information. OVE noted during this validation cycle that, although in line with the guidelines, XSRs calculate the achievement of objectives as the proportion of value achieved over target, not considering the baseline of the value.²⁰

2.16 Half the operations were rated efficient. A positive efficiency rating means that a project's financial and economic benefits outweighed its costs and that it delivered on the process and business objectives stated at approval. Of the 28 operations with FIs, 13 had a negative rating. Of these, nine exhibited insufficient or even negative growth and performance of the relevant portfolio—generally micro, small, and medium-sized

17 In this case, low relevance was also related to the choice of instrument that turned out to be limited.

18 A negative effectiveness rating does not necessarily mean that the project did not achieve any of its development objectives, but that it had an objective rated unsatisfactory or that it only met half or less of its objectives.

19 Six of these cases were operations with FIs that aimed at increasing their SMEs portfolio and tenors. At end of project, portfolios contracted and, when applicable, tenors shortened. Three other cases include corporate investments aimed at increasing local employment and purchases from local suppliers. At end of project, the associated indicators showed a decline.

20 PCRs calculate achievement ratio as the proportion of value achieved over target relative to their baseline (*PCR Guidelines* 2020, [OP-1696-5](#)).

companies. For corporate and infrastructure and energy operations, efficiency is assessed with financial and economic indicators. Of eight projects, five had a negative efficiency rating, three of which were due to low or even negative returns on invested capital (ROIC). The other two cases had a negative financial rate of return (FRR) and an economic rate of return (ERR) lower than the company's weighted average cost of capital (WACC).

- 2.17 The results in just over half the operations (53%) were likely to be sustainable. Seventeen operations had negative sustainability ratings—eight of them were FI operations that failed to expand their target portfolio and exhibited a deteriorating portfolio performance. Although less frequently, the sustainability of results was also hindered by risks related to the operating environment²¹ (three cases), followed by low financial indicators of clients (two cases) with drops in their risk rating and increases in non-performing loans.²²
- 2.18 In environmental and social performance, one operation generated negative environmental impacts. The operation released low-quality industrial wastewater, causing concentrations of pollutants that exceeded local criteria giving rise to community complaints. In addition, the client of an FI operation did not ensure that all sub-borrowers were compliant with IDB and country-level laws and regulations.

3. Non-core criteria: Investment outcome, work quality, and additionality

- 2.19 Most operations proved profitable for the IDB Group, with 86% rated positive in their investment outcomes. Almost all operations reviewed corresponded to loans (35 of the 36).²³ Loan profitability is measured in terms of the proceeds from the loan and whether they will occur as projected at the time of approval. The negative ratings are mainly explained by prepaid projects (four operations of the five with negative ratings) and one partially canceled project. In these cases, IDB Invest expects to receive less than 60% of the projected interest payments.

21 For example, in one case the project was vulnerable to changes in the unit cost of electricity and to variations in sunlight and transmission availability. While only the first identified risk materialized, and a mitigation measure put in place, the project's risky profile affected the sustainability rating. Another case was a company operating in a highly protected sector that saw the sustainability of its long-term results at risk due to the change in trade policy (lifting tariffs) and subsequent increased competition. Finally, another case was a public bank with strong linkages to the macroeconomic situation of the country and consequently to the sovereign's rating behavior.

22 Other less frequent cases of low on sustainability rating relate to: (i) low ratings on E&S performance (¶2.20); (ii) contraction of FI portfolio along with the annual growth of the relevant SME portfolio; and (iii) lack of information to assess project sustainability.

23 The other operation was equity.

2.20 Only two of the 36 operations reviewed had a positive rating for work quality. It is also the criterion with the lowest rating during this cycle. This criterion analyzes the quality of the IDB Group's work around two aspects: (i) selection, analysis, and structuring of the operation at design, and (ii) monitoring and supervision of the operation during execution. Most operations (34 of the 36 analyzed) had a negative rating. Of these, 24 suffered from design shortfalls (i.e., ill-defined or inappropriate objectives, indicators, and targets; limited diagnoses; and suboptimal financial analyses), that is mostly M&E issues at entry, and only two presented deficiencies in supervision. Eight cases presented flaws in both aspects.

2.21 Three-quarters of operations (75%) showed additionality. This criterion seeks to estimate the additionality, or added value, that IDB Group support brings to each project. Both financial additionality (e.g., whether the IDB Group granted financial conditions unavailable to the client at the time of approval, or if it mobilized additional resources) and nonfinancial additionality (e.g., whether the IDB Group improved the project design, or if it contributed to a better performance of the customer's operation) are analyzed. While only one of the negative ratings stemmed from evidence of inexistent financial additionality,²⁴ most negative ratings (8 of the 9 cases) are explained by insufficient evidence of the project's financial additionality. Projects are supposed to be evaluated using a counterfactual assessment of how the project would (or would not) have proceeded without IDB Group support. But most XSRs assessed the client's liabilities. They did not examine, for example, the tenors available from other sources at the time of approval. For the cases where nonfinancial additionality was expected, low ratings were frequently related to technical cooperation components that did not materialize and were supposed to improve client's internal performance.

4. XSR quality

2.22 Most XSRs (92%) were good quality. Although no XSR was rated excellent during this cycle, OVE rated 33 out of the 36 of them as 'good' quality. For these cases, XSR guidelines were applied in most respects, data was sufficient and clearly presented, project results were generally captured and reported in a balanced way, lessons were based on clear evidence, and only minor shortcomings in clarity, consistency, and candor were found. The three evaluations rated fair referred to XSRs

24 The loan was supposed to provide a tenor and grace period not available at the local markets. This assumption was challenged in 2017, however, when the client prepaid the loan and declined the remaining proceeds claiming that the transaction implied higher costs compared with financing opportunities in the local markets.

for FI operations with critical shortcomings on some evaluation criteria, identification of the relevant portfolio, and related indicators.

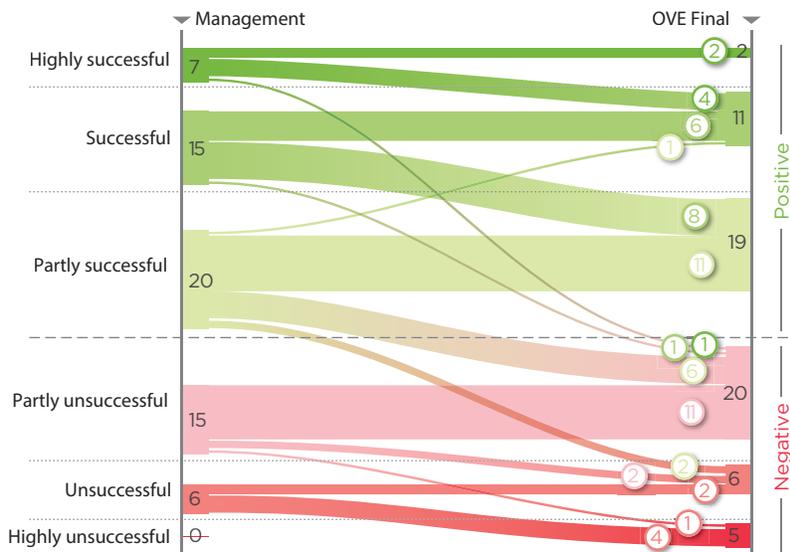
5. Comparison of OVE and Management’s ratings in PCRs and XSRs

2.23 Of the PCRs with comparable scores (62 of 63), OVE and Management had the same ratings on overall project outcome for about half the operations (32) (Figure 2.4). OVE downgraded the overall outcome ratings for 29 SG operations rated higher by Management and upgraded only one. A third of the discrepant overall outcome ratings (10 operations) occurred with positive Management ratings but negative OVE ratings. Although most of these operations (eight out of 10) were on the threshold of a negative rating, two of them represent the most critical discrepancies: BR-L1252 and PN-L1151 had 3- and 2-point rating downgrades on overall outcome, respectively.²⁵ Most downward adjustments on overall outcome (16) are explained by changes in one of the core criteria (relevance, effectiveness, efficiency, or sustainability). Relevance was downgraded most frequently (poor assessment of country realities, weak vertical logic), followed by effectiveness (low-quality indicators in the results matrix, lack of information on certain indicators at project completion, or ill-defined indicators for certain objectives).

Figure 2.4

Number of SG operations rated in each overall project outcome category by Management and OVE, and changes among these categories

Source: OVE.



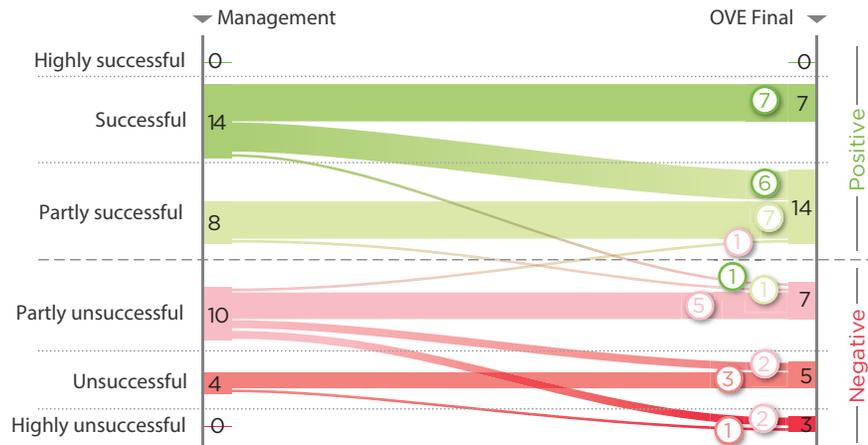
25 OVE’s adjusted ratings for effectiveness impacted the overall rating for outcome. The project BR-L1252 did not have properly defined outcome (but output) indicators to measure progress. The series PN-L1151 lacked information of indicators to measure progress. The PCR team did not apply the guidelines for PBPs correctly. The latter require that achievement is measured with respect to the overall success at end the program.

2.24 On the NSG side, OVE and Management assigned the same overall project outcome ratings in 22 of the 36 cases.²⁶ OVE provided a different overall project outcome rating in 14 of the 36 operations evaluated (39%). OVE’s rating was generally lower than Management’s, with 13 downgrades out of the 14 cases with discrepancies. From the cases where OVE and Management differed on overall project outcome rating, two are the most critical because Management rated them positively while OVE’s downgrades produced a negative rating. One of these cases was rated negative because the achievement level in one objective was lower than the baseline. The other case had downgrades in effectiveness, efficiency, and sustainability. Of the 13 downgrades, 10 were downgraded by one point. In eight cases, OVE’s assessment differed from Management’s in one of the four core criteria.²⁷ Downgrades were concentrated on the effectiveness criterion, followed by sustainability, efficiency, and relevance. Effectiveness downgrades were related mostly to unmet targets (five cases) or insufficient information to account for progress (two cases).²⁸

Figure 2.5

Number of NSG operations rated in each overall project outcome category by Management and OVE, and changes among these categories

Source: OVE.



26 Management rated 61% of the operations positive, while OVE rated 58% positive.

27 In four cases, OVE’s assessment was different from Management’s for two or three, and in one case the assessment differed for all four core criteria.

28 Other cases with downgrades in effectiveness are explained by separating objectives and different interpretations of guidelines from the XSR team.



03

Projects
Reviewed by
OVE, 2017-2020

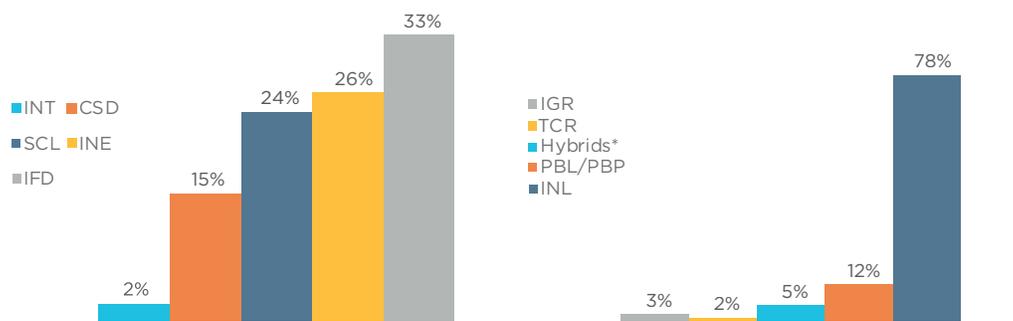
3.1 To date, after four validation cycles in accordance with the unified objectives-based methodology, OVE has completed 183 validations of SG operations and 144 validations of NSG operations. This allows to take stock of validation results across multiple years to analyze trends and discern lessons. This chapter reviews the results of OVE’s validations between 2017-2020 to the extent that such results are comparable and undertakes a deep dive into the lessons drawn by the self-evaluations.²⁹ Given that the composition of projects assessed changes from year to year, particularly for NSG, the following section focuses on overall results between 2017-2020 rather than on assessing trends or changes between years.

A. Public sector operations (SG)

3.2 The distribution of operations by sectors and instruments was relatively homogeneous among the four validation cycles carried out by OVE (2017–2020), with the greatest number coming from Institutions for Development (IFD) (Figure 3.1). Between 2017–2020, OVE reviewed and validated 183 SG operations. Most of them (80%) consisted of operations approved before 2013. The biggest proportion of operations came from the Institutions for Development (IFD), followed by Infrastructure and Environment (INE), Social (SCL), and Climate Change and Sustainable Development (CSD) respectively. Only 2% were from Integration and Trade (INT). In terms of instruments, the majority corresponds to investment loans (143, or 78%). Policy-based Loans (PBL) and Policy-based Programmatic Loans (PBP) comprised 12% (22) of the operations.

Figure 3.1
Distribution of validated operations by departments and instruments†, 2017–2020

Source: OVE.



† TCR corresponds to reimbursable technical cooperation

* Includes INL+IGR; INL+PBL; and PBL+TCR.

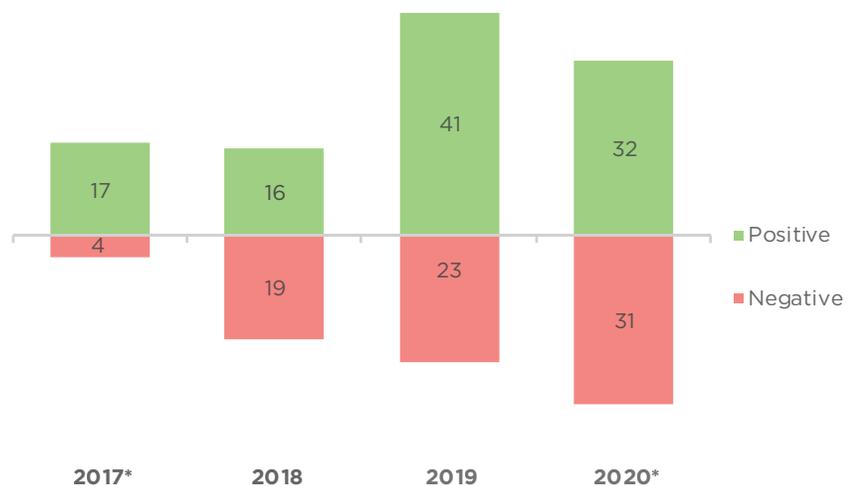
29 On the SG side, while the PCR guidelines changed in 2018, in agreement with Management, OVE had already applied the methodology contained in the 2018 guidelines in its 2017 validations, thus OVE’s validation results from 2017 through 2020 are comparable. On the NSG side, validation results for the effectiveness criterion and the overall outcome rating are only comparable for the validation cycles 2018–2020 due to adjustments in the methodology in 2017–18. For this reason, ratings for overall outcome and effectiveness are assessed only for the validation cycles of 2018 through 2020.

1. Overall outcome ratings, 2017–2020

3.3 In total, 106 of the validated operations, or 58%, achieved a positive overall outcome rating.³⁰ By sector, IFD had the highest proportion (31%) of positive overall outcome ratings, followed by INE (30%). In terms of lending instrument, while 57% of investment loans achieved a positive rating, this proportion was 64% for PBLs and PBP. Conversely, 43% (77) of validated operations had a negative overall outcome rating.³¹ Most operations with negative ratings for overall outcome (82%) had at least 2 core criteria rated negative — most frequently effectiveness and efficiency. All operations with a negative overall outcome rating, however, had a negative rating for effectiveness.³² Overall outcome ratings are presented in Figure 3.2.

Figure 3.2
Ratings on overall project outcome by validation year (number of operations)

Source: OVE.



*Includes projects with CO before 2016.

2. Core criteria

3.4 Relevance shows the highest rating by criterion, with 92% of operations rated positive. These results reflect the alignment of IDB operations with country development challenges, IDB country strategies and corporate goals, and an adequate vertical logic of operations at design phase. This year's relevance ratings were on average the lowest. Although relevance was positive in at least 95% of reviewed operations during previous validation cycles (2017 through 2019), it was 84% in the 2020 cycle. The decline is explained by a few projects that were poorly

30 Of the total 183 validated SG operations, 3% (5) had a highly successful rating, while 22% (40) were rated successful.

31 Three percent (6) had a highly unsuccessful rating, and 9% (17) had an unsuccessful rating.

32 Negative effectiveness ratings were driven by a combination of factors: the quality of the results matrix, changes to the original results matrix after the allowed time, and cancellation of components or products.

aligned with country realities, a shortcoming that affected implementation or resulted in canceled components or funds that distorted the vertical logic of the program.³³

- 3.5 Overall, only 32% (59) of operations garnered positive ratings³⁴ for effectiveness, making it the criterion with the lowest ratings.³⁵ A review of the 183 completed validations found that a combination of factors explains negative effectiveness ratings, but poor M&E emerges as a key factor.³⁶ Negative ratings for effectiveness appear to have arisen from (i) poor quality results matrix or lack of M&E compliance (87 cases);³⁷ (ii) underperformance of projects (56 cases), explained by risk materialization (i.e., political or institutional changes, challenges to institutional coordination, or macroeconomic or social crises),³⁸ lack of diagnosis of the context/relevant target population,³⁹ or an overambitious design;⁴⁰ and (iii) cancellation of products or components that prevented achievement of outcomes (37 cases).⁴¹

33 For example, the objective of the project BH-L1030 was “to promote human capital accumulation and poverty alleviation, through the consolidation of existing programs into a Conditional Cash Transfer (CCT)”. Although the loan proposal identified some risks, such as coordination challenges between participating Ministries, it failed at identifying the risks that would bring a change of government, and at overestimating technical capacities PCR, BH-L1030). These overlooked risks materialized, resulting in the cancellation of 45% of original approved resources. The CCT component was not delivered.

34 By sector, the most positive ratings for effectiveness belong to IFD (30%) and INE (30%). In terms of instrument, while 29% of investment loans had a positive rating for this criterion, the proportion for reform programs was 39%.

35 Systematically, the effectiveness rating has been the lowest rated criterion across validation cycles. It achieved its highest value in the 2017 cycle with 52% PCRs rated positive, and its lowest in 2018 with 17% PCRs rated positive.

36 The three main factors leading to negative effectiveness are not mutually exclusive, as some projects were affected by multiple factors.

37 M&E issues included problems at entry (i.e., results matrix included indicators that measured outputs instead of outcomes; were overbroad or irrelevant to the project-specific objective; showed no baselines or targets; lacked attribution; or results matrix did not include an indicator for a specific objective), and during implementation (i.e., indicators were not followed up; omission of original indicators or indicators introduced belatedly; challenges related to compliance or capacity to carry out the M&E plan).

38 For example, the project BR-L1287 faced challenges that affected delivery of results: complex and delayed negotiations between the executing unit and the state secretaries, social and fiscal crisis in the state, and constant institutional changes (i.e., a change of at least 6 secretaries and 4 coordinators).

39 For example, the project BR-L1442 had no progress on one of its outcome indicators because lack of demand meant it financed no energy efficiency projects. The PCR notes that future operations could benefit from a more thorough analysis on country context and demand.

40 For the project NI-L1067, the original execution scheme was inefficient and required redesign and simplification. The initial scheme resulted in significant delays and implementation difficulties in the first two years of the program (i.e., at half the time allotted for execution, 25% of the resources had been disbursed and only 18% executed).

41 Eighty-four percent of operations negatively rated on effectiveness presented at least two of the issues discussed in this paragraph.

- 3.6 Around half (53%) of IDB investment loans reviewed proved cost-effective or cost-efficient. Most assessments employ a cost-benefit (CBA) or cost-efficiency (CEA) methodology that includes the direct outcomes and/or costs of the project rather than a cost-and time-overrun analysis (CTOA) based on PMR ratings, the Cost Performance Index (CPI), and the Schedule Performance Index (SPI) during implementation.⁴² The use of these methods and their quality improved across validation cycles, which, in part, explains the better ratings.⁴³ Most operations (63%) with negative efficiency ratings employed either a CTOA for the efficiency analysis or a comparison between planned and executed budgets. The rest of them had either a low-quality CBA or CEA (29%) (i.e., not credible assumptions or comparators, attribution problems for results, or partial analysis) or turned out to be inefficient (8%).
- 3.7 Sustainability is the criterion with the second highest positive ratings (71%). This criterion includes two aspects: the probability of continuation of outcomes, and safeguards performance. Although most PCRs contained a systematic analysis of risks to continuation of outcomes, the information about safeguards performance or the implementation of environmental compliance plans was often inadequate for category B operations.⁴⁴ Most negative ratings on sustainability stemmed from poor (or absent) mitigation measures to secure continued outcomes (e.g., 84% of these projects struggled to secure funds to maintain or operate project infrastructure or outputs).

B. Private sector operations (NSG)

- 3.8 Portfolio composition has changed throughout validation cycles. Operations with FIs have increased. In line with the IDB Group's institutional changes—i.e., the consolidation of the private sector windows into IDB Invest—the relative share of the SCF portfolio dropped over the review period. The proportion of former IIC operations increased and represents more than half the overall reviewed portfolio. By segment type, the proportion of corporate and infrastructure and energy operations has decreased over time. Considering all years, corporate and infrastructure and energy operations represent 39% of the portfolio, while projects with FIs accumulate the remaining 61% (Figures 3.3 and 3.4).

42 Note that these instruments consider only output progress, and automatically lead to a negative efficiency rating.

43 Before 2018, this criterion could achieve a positive rating with a CTOA analysis or a comparison between planned and executed budget. This difference affected three of the 21 operations from the 2017 cycle.

44 Of the reviewed operations, 35% were category B, and only 1 program was category A.

Figure 3.3

Evolution of NSG portfolio by window

Source: OVE, with data from Analitika's "Maestro All Transactions" Report, XSR reports, and Enterprise Data Warehouse.

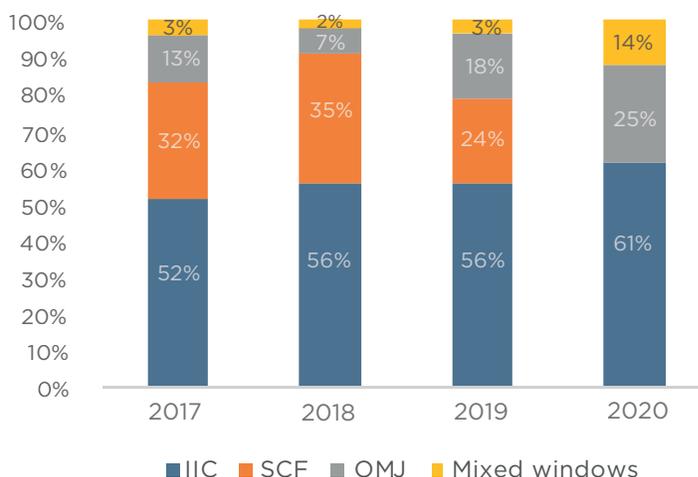
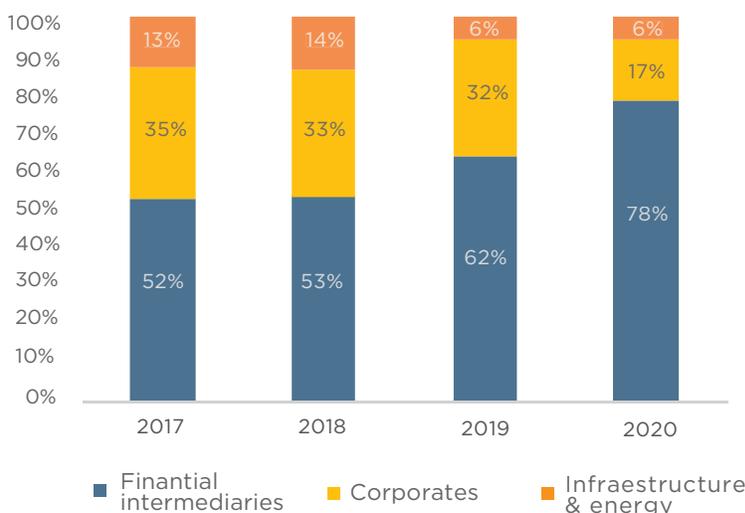


Figure 3.4

Evolution of NSG portfolio by segment

Source: OVE, with data from Analitika's "Maestro All Transactions" Report, XSR reports, and Enterprise Data Warehouse.



1. Overall project outcome ratings, 2018-2020⁴⁵

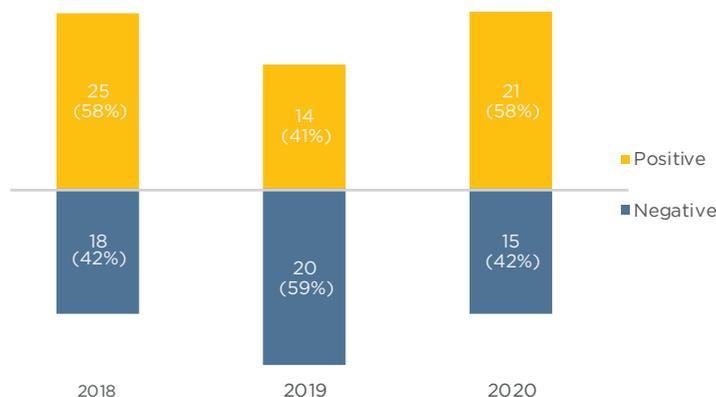
3.9 In total, 53% of operations had positive overall project outcome ratings. Ratings in 2020 are higher than they were for the 2019 cycle, driven by improvements on all 4 core criteria. Any interannual performance differences must be interpreted with caution, however, given the scant number of operations evaluated each year. The 2018-2020 validation cycles involve projects approved between 2005 and 2015. During this period, important changes were introduced with the aim of improving the development effectiveness tools and standards for project appraisal, approval, and supervision (Box 3.1). An important milestone was the call in 2013 for a results matrix for IDB's NSG operations.

45 Overall outcome ratings are comparable only for validation cycles that implemented the 2018 XSR guidelines. Prior to this cycle, 2015 pilot guidelines were used. For these cases, effectiveness was rated differently because it did not include the indicators-achievement thresholds to assign the rating. In addition, the pilot guidelines did not specify that that pure financial and business objectives (e.g., improving asset liability matching) did not constitute valid development objectives, nor did it include the FI interim guidelines that clarified the assessment of the FI relevant portfolio.

Figure 3.5

NSG overall project outcome ratings by validation year (number and percentage of operations)

Source: OVE



Recuadro 3.1. Herramientas de efectividad en el desarrollo para operaciones del sector privado

The IDB launched the Development Effectiveness Framework (DEF) in 2008 (GN-2489), introducing the Development Effectiveness Matrix (DEM). In the same year, SCF developed its own framework (GN-2473-1) with an NSG DEM that introduced development outcomes and ‘additionality’ as key criteria at the initial stage of project screening and selection. As such, all SCF operations approved after March 2008 would have a DEM on their loan and guarantee proposals. OMJ adopted SFC’s DEF with some adjustments to reflect a project’s specific characteristics. Several changes were introduced to the DEM in 2011 per GCI-9 agreement, including the evaluability score and a minimum threshold for approval. In 2013, the NSG DEM was revised (GN-2489-8). It incorporated a results matrix to the development effectiveness toolkit. Hence, each NSG loan proposal would include a results matrix stating the project objective, its output and outcome indicators.

The IIC had its own system to track development results and ex-ante assessment was conducted through the Development Impact Assessment Scorecard (DIAS). The DIAS was introduced in 2008 and provided a score for development outcome and additionality. Later on, the DIAS Plus was introduced. After the 2016 merge-out of the IDBG’s private sector operations into the IIC, the Development Effectiveness Learning, Tracking, and Assessment tool (DELTA) became the system for ex-ante scoring of development results and additionality of operations with the private sector.

2. Core criteria: Relevance, effectiveness, efficiency, and sustainability

3.10 Two-thirds of operations validated between 2017 and 2020 were relevant. Among those with negative ratings, design issues were the most frequent reason (85%),⁴⁶ including problems related to

⁴⁶ As corroborative evidence, OVE assessed the work-quality ratings from 2017 to 2020. Negative work-quality ratings emerge most frequently from precommitment issues (screening, appraisal, and structuring). Of cases with negative work-quality ratings, 56% stemmed exclusively from these problems, compared with 2% arising exclusively from monitoring and supervision flaws and 41% presenting both issues.

weak diagnostics, inadequate or unclear mitigation, unjustified choice of client, inadequate loan size, lack of baselines, missing or inadequate targets. Progress has been made regarding alignment of operations with IDB Group's priorities and development needs. Likewise, progress has been made on having fewer operations in distorted environments.⁴⁷ On average, operations with FIs had lower relevance ratings than corporate and infrastructure and energy operations (61% of FI operations had positive relevance ratings, compared to 74% for corporates and 79% for infrastructure and energy), although they saw improvements during the last validation cycle, reversing the trend seen between 2017 and 2019.

- 3.11 Of operations reviewed between 2018 and 2020, 45% were effective. The negative effectiveness ratings are explained by insufficient information to account for progress and by underachievement of targets, with the latter being more frequent.⁴⁸ For a handful of cases (5) with unmet targets, XSRs explain that underachievement stemmed from problems with the results matrix (e.g., overambitious targets, lack of baselines, inappropriate indicators). For the rest, external factors (48% of cases) was the most frequent explanation or no clear reason is provided (37% of cases).⁴⁹
- 3.12 Although efficiency has improved, it remains the lowest-rated criterion; 40% of operations were rated positive between 2017 and 2020. Improvements on efficiency were driven mostly by FI operations,⁵⁰ which have progressed on collecting data on the target portfolio's performance.⁵¹ Underachievement of FI portfolio targets, the quality of portfolios (non-performing loans — NPLs), and low financial and economic returns continue to adversely affect efficiency ratings.
- 3.13 Between 2017-2020 about half the operations were sustainable, with higher ratings traced to better environmental and social performances and lower ratings related to the continuation of results. Fifty-three percent of FI operations were rated positive, while those with negative ratings cited risks to the continuation of results.⁵² Corporate operations had higher ratings (62% positive;

47 For example, supporting a public company with a dominant position in the market, or supporting access to finance in contexts where financial deepening is limited due to regulations on interest rates and other financial and capital market aspects.

48 Of the 62 operations with negative ratings, 47% had unmet targets, 18% insufficient information, and 35% cited both.

49 The other 15% of cases cited internal factors or a combination of both internal and external factors.

50 The proportion of FI operations rated positive on efficiency has increased from 12% in 2017 to 54% in 2020.

51 The proportion of FI operations reporting performance of target portfolio has increased from about half of cases to more than 90%. Performance of the target portfolio (e.g. growth, non-performing loans, and average interest rates) is used as a proxy to assess the contribution of the target portfolio to the FI's profitability.

52 Among FI operations, 86% had a positive rating on E&S performance.

26 out of 42),⁵³ while infrastructure and energy operations were 36% positive (five out of 14); those with low sustainability cite both E&S performance and continuation of outcomes. Assessments of sustainability have changed over time, so that mere compliance with E&S standards will not automatically produce a positive rating. As per XSR guidelines, high probability of continuation of results is also required.⁵⁴

3. Comparisons of OVE and Management ratings in PCRs and XSRs

3.14 IDB Management and OVE have generated similar ratings discrepancies for overall project outcomes over the past two years. A key purpose of the independent evaluation office's validation of project outcome ratings is to ensure the credibility of the self-evaluation system. Therefore, it is useful to compare OVE's final project outcome ratings with Management's own ratings. Methodological differences make comparisons impossible for PCRs prepared in accordance with 2014 guidelines. Therefore, only PCRs reviewed in the 2019 and 2020 cycles are comparable, excluding those PCRs still prepared according to 2014 guidelines from the 2019 and 2020 batches.⁵⁵ For comparable PCRs, the number of Management's self-evaluations ratings that differed from OVE's overall project outcome ratings increased from 46% (25 out of 54 cases) in 2019 to 48% (30 out of 62 cases) in 2020. Most differences resulted in a lower overall project outcome rating from OVE. Overall, while OVE found that 56% of SG operations achieved an overall positive outcome rating, Management assigned a positive rating to 75% of operations. Most discrepancies are explained by downgrades in 1 (27% of cases) or 2 criteria (45% of cases). The criteria most frequently downgraded were effectiveness and efficiency. While the former was due to the quality of the results matrix and missing information from original indicators, the efficiency criterion was affected by the quality of the CBA or CEA analysis.

3.15 For NSG operations, discrepancies for overall project outcome ratings decreased between 2018 and 2020. In 2018, OVE provided a different overall project outcome rating for 60% of operations, a proportion that fell to 39% in 2020.⁵⁶ From the cases with different ratings across all three years (51), 25% (13 cases) drew positive ratings from Management but negative re-ratings from OVE. This

53 Fifty-eight percent of corporate and infrastructure and energy were rated positive on sustainability.

54 In the validation cycles of 2017 and 2018, it was possible to find operations with a positive sustainability rating, driven by compliance with E&S standards (not exceptional E&S performance).

55 The 2019 validation cycle included 10 of the 64 PCRs prepared under 2014 Guidelines, and the 2020 validation cycle included 1 of the 63 PCRs. Thus, comparisons between self-ratings and OVE ratings are possible only for 54 operations from the 2019 cycle and 62 from the 2020 cycle.

56 In total, considering all three years, OVE provided an overall project outcome rating different from Management's in 44% of cases (51 operations).

type of discrepancy also decreased from 21% in 2018 (nine out of 43 cases) to 6% in 2020 (two out of 36 cases). In these cases, OVE downgraded two, three, and even four of the core criteria. Differences in effectiveness ratings were most frequent, with downgrades in the achievement of one or more development objectives to unsatisfactory (11 of the 13 cases). Efficiency and relevance followed in downgrade frequency (nine of the 13 operations were downgraded on efficiency, and eight on relevance).



04

Lessons Learned

4.1 OVE analyzed the lessons included in all the PCRs and XSRs validated between 2017 and 2020. Institutional learning is key for the IDBG's development effectiveness as it can play a major role in improving results in future operations. PCRs and XSRs are meant to work as tools for both accountability and learning. The lessons included in these reports often emerge from actions that proved successful. In other cases, they emerge from experiences that led to disappointing outcomes, and that were considered important to avoid in the future. In this section of the report, OVE presents lessons in all validated PCRs and XSRs between 2017 and 2020.⁵⁷

A. Review of lessons

4.2 In an effort to capture their main topics, OVE categorized the lessons from PCRs and XSRs. Content analysis techniques were used to summarize, code, and then classify lessons into concise categories.⁵⁸ OVE determined the core message by identifying each lesson's specific call to action. For example, two SG operations stated the following:

Currently, the executing unit works as a coordinator of other units, such as Planning, Finance, Procurement, Legal, etc. This often caused delays or made the implementation more difficult, as the other units do not have the human resources with the necessary capacities to deliver to the extent required. To strengthen the executing unit, the [project] should include capacity building for all officers involved in technical areas, finance, legal, procurement, etc., to contribute to improve the executing unit's implementation capacity (ES-L1045).

The executing unit's weaknesses compromise the implementation capacity of the project. It is necessary to strengthen the institutional capacities of project executors by creating units of excellence [...]. [Future projects] should create their executing units with exclusive dedication, with officials from the permanent staff to guarantee the continuity of the projects and the internalization of knowledge [...] [and also] incentivize the constant evaluation of the executing unit's institutional capacity, by detecting deficiencies and creating action plans in response to correct them (BR-L1252).

57 Unlike XSRs, PCRs do not have a section titled 'lessons learned' per se, but the findings and recommendations section is supposed to include any lessons.

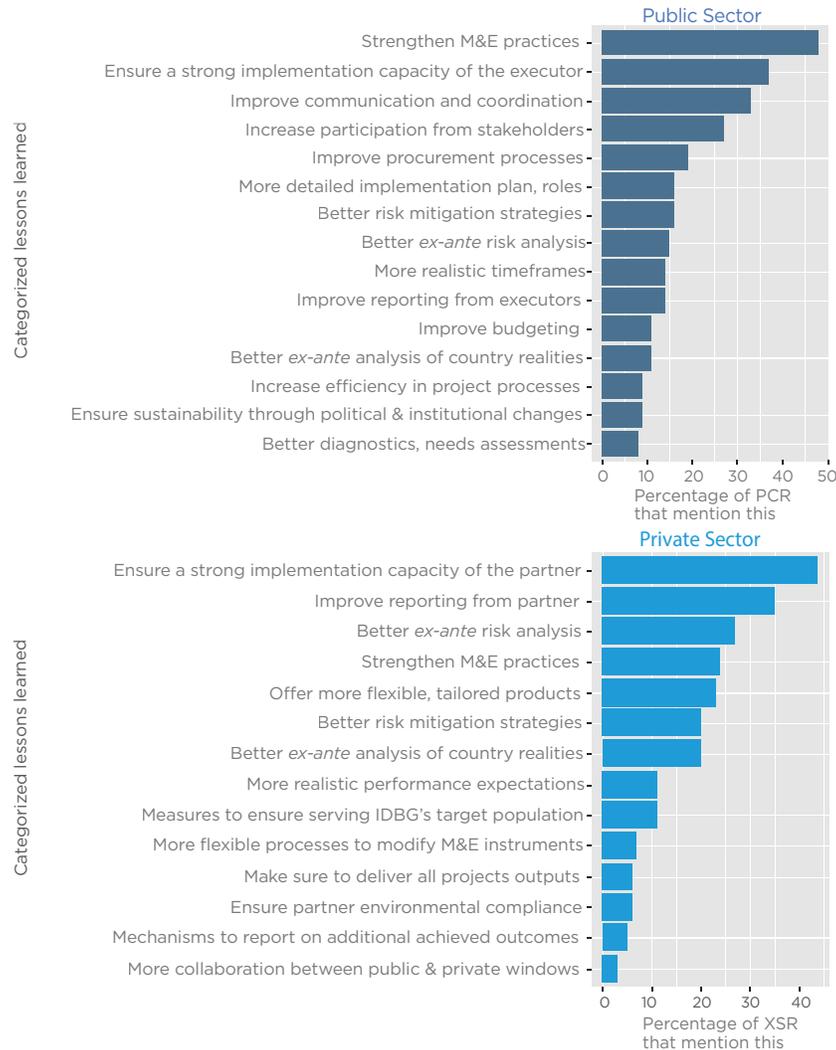
58 This analysis was entirely inductive (as opposed to deductive) in the sense that codes and categories were not defined ex-ante, but rather emerged directly from the raw lessons data as OVE staff reviewed the text for each lesson.

As the call to action here is to ensure that the executor has a solid implementation capacity, this type of lesson was categorized as “Ensure a strong implementation capacity of the executor.” Several rounds of reviews were conducted to ensure that lessons had consistent categorizations and produced a final and valid set of categories. Figure 4.1 summarizes the results of this content analysis, on which the rest of the findings in this section are based.

Figure 4.1

Categorized lessons learned from PCRs and XSRs

Source: OVE’s categorization; lessons learned text data partly extracted from the PCR [Lessons Learned Dashboard](#) and completed by extracting the text from the full PCR reports, and the full XSR reports as well.



Note: Percentages out of a total of 177 PCRs and 141 XSRs reviewed. Percentages do not add up to 100% as categories are not mutually exclusive. For PCRs, only the top 15 categories are plotted above out of 27 identified. The remaining ones are: offer more tailored products (mentioned by 8% of PCRs), more realistic targets (6%), allow time to measure outcomes and impact (5%), more flexible processes to modify M&E instruments (5%), engage the executing unit throughout all project stages (3%), mechanisms to report on additional achieved outcomes (2%), and 6 more categories with only 1% or less. Only seven PCRs had lessons that did not fit into any of these categories. For XSRs, all 14 identified categories are plotted. Only four XSRs had lessons that fell outside these categories.

4.3 The most frequent message in PCRs, and the fourth most frequent in XSRs, was to improve M&E practices. As shown in Figure 4.1, an important share of the lessons from both PCRs and XSRs focus on strengthening M&E practices in the operations.

Examples include the need to establish clear project objectives, design high-quality indicators with appropriate targets, obtain baseline data for all indicators, measure the indicators consistently, comply with the SMART criteria (specific, measurable, achievable, relevant, and time-bound indicators), and specifying their calculation methods and data sources. These lessons also stressed to avoid subjective and costly indicators, those not able to measure project outcomes correctly, or those that were obtained from unsteady data sources. Some also emphasized the need to use a smaller number of indicators, measuring baseline values for all of them, setting clear dates for updating M&E instruments, and defining a clearer relationship between outputs, outcomes, and impacts.

- 4.4 A text analysis of the lessons confirms the focus on M&E and calls for a revision of the current data collection mechanisms. To complement the categorization from Figure 4.1, OVE used text mining techniques to further analyze the ‘lessons learned’ sections of PCRs and XSRs (see detailed results in Annex VII). It found that issues related to ‘indicators’ and ‘information’ are equally predominant in lessons from both SG and NSG operations, thus confirming previous findings. Moreover, the contexts in which the word ‘indicators’ was used (see Annex VII) reveal that there were expectations that indicators failed to meet, generally because they were not carefully designed at first and their flaws only became apparent during project implementation. At times, inconsistencies in the data provided by clients compromised its reliability. In some cases, clients’ failure to comply with reporting requirements prevented the IDBG from being able to use that data to properly calculate the indicators. As for the word ‘information,’ the contexts (see Annex VII) show frequent problems with data sources, lack or poor availability of information, as well as deficiencies in data quality and issues with collecting relevant information from clients. Overall, the frequent recurrence of these lessons across public- and private-lending windows alike reveals much room for improvement in the application of good M&E practices — at the IDBG and among its clients. Conclusions also point to a need for a careful revision of the mechanisms currently used for data collection. Many operations are unable to retrieve the information they need to properly calculate the indicators for measuring project performance, leaving a large knowledge gap on whether the IDBG’s operations in the region are achieving the expected results.
- 4.5 Content analysis revealed that the second most frequent category for PCR lessons, and the first for XSRs, was to ensure a strong capacity for implementation on the executor’s or partner’s side (Figure 4.1). In SG projects, this capacity generally

referred to building a strong team for the executing unit. The lessons point to the need for executors with substantive sector experience, strong technical and operational capacity, and proven financial management ability. In cases where these were lacking, the lessons emphasized the need to strengthen the executor's capacities through technical cooperation, training, or other means. In NSG projects, this category of lessons typically advised choosing a partner that is financially strong and has ample experience with the target segment or ultimate beneficiaries, with objectives aligned with those of the IDBG and a record of effective risk-mitigation in difficult economic times. Some lessons in this category also emphasized that the most adequate partner might not always be the strongest financially but the one most committed to the IDBG's development objectives. In these cases, the lessons suggest providing technical cooperation and capacity building, to delegate challenging tasks to an experienced third party, or to choose partners that are backed by a strong parent company, sponsor, or guarantor.

- 4.6 Another frequent exhortation in PCRs was to improve communication, coordination, and participation among relevant stakeholders. As shown in Figure 4.1, many PCRs called for sound communication and coordination among the agents involved in the project. In some projects, this led to highly positive results; where these were lacking, projects underperformed. Another set of lessons relate to the positive effects of participatory approaches to projects. For example, in one education project, engaging parents helped to secure school improvements. Conducting participatory processes with citizens and local authorities, such as public consultations and community engagement activities, proved effective for understanding the needs of the population and adapting the project to local needs. Positive, sustainable outcomes were also observed when participation was encouraged from other government offices, subnational governments, private companies, nonprofits, academia, and other relevant entities within the sector.
- 4.7 Lessons learned have focused more on how to improve project design, and less on how to ensure an effective implementation. Many of the lessons cite poor project design, having observed the occurrence of several issues that could have been anticipated and then mitigated during project implementation. Fewer lessons mention concrete actions taken during implementation that made a project more resilient so it could reach its goals despite those issues arising. When goals are not achieved, it is typically reasoned that risks that were not seen as major threats at design ended up occurring during implementation.

B. Institutional learning

4.8 IDBG has developed two key platforms for improving institutional learning processes. IDB launched the [PCR Lessons Learned Dashboard](#), where lessons from previous projects are collected. They can be filtered by department, sector, year, and other variables. While users can read the data directly in the platform, they cannot download it entirely to add notes, manipulate it in any way, or merge it with other relevant data for their projects, all of which could be useful to permit analysis of the lessons.⁵⁹ IDB Invest also launched the [Development Effectiveness Analytics](#) (DEA) platform, where staff can access lessons from completed projects.⁶⁰ (The restricted download on the IDB side notwithstanding, these platforms are major steps toward an institutional culture that takes learning seriously and that is committed to improving results.

59 In the case of IDB, full access to the entire lessons data in a downloadable format requires a formal request to the platform administrators, which is assessed on a case-by-case basis.

60 Full access to the lessons learned dataset will be available for IDB Invest staff in 2021. Other IDBG staff may obtain access to the platform by submitting a formal request.



05

Conclusions and Recommendations

- 5.1 In 2020, OVE conducted 99 validations—63 of PCRs and 36 of XSRs. While 51% of SG operations achieved an overall positive project outcome rating, the proportion was 58% for NSG operations. As in previous cycles, effectiveness ratings were the lowest for both SG (29% positive) and NSG (50% positive) operations, driving overall development outcomes down. For SG operations, negative effectiveness ratings are mostly explained by a combination of factors: poor quality M&E, poor performance and cancellation of products that prevented achievement of outcomes. NSG operations, particularly those with FIs, were influenced mostly by unmet objectives, which led to negative effectiveness ratings. Their poor results in relation to the expansion and performance of their target portfolio also led to low efficiency ratings and results were unlikely to be sustainable. Regarding reasons for unmet objectives, XSRs frequently hypothesize that failure to reach objectives was related to factors beyond the control of the client (e.g., economic crises, changes in regulatory environment, decreased demand for SME credits), but they rarely provide a clear analysis of how project design and/or supervision fell short on taking such external factors into account.
- 5.2 Over the last four years, OVE has reviewed and validated the performance of 183 SG operations. A systematic analysis from all validation findings suggests that low-quality M&E affects a large share of operations and is a key contributor to low overall project outcome ratings. Of all SG validated operations, 53% achieved an overall positive outcome rating. Most operations with overall negative outcome ratings, had at least 2 core criteria rated negative—most frequently effectiveness and efficiency. Negative effectiveness ratings arose from a combination of factors ranging from poor M&E quality, underperformance of projects (risk materialization—political or institutional changes, challenges for institutional coordination, or macroeconomic or social crises); cancellations of outputs that prevented achievement of outcomes; lack of diagnosis regarding context or relevant target population; and overambitious design.
- 5.3 Cancellation of components or outputs for SG operations explain a number of negative effectiveness ratings. The systematic analysis from all validations reveals that partial or complete cancellation of components or outputs linked to outcomes indicators altered the vertical logic of projects and culminated in negative effectiveness ratings. Cancellations materialized in almost a third of projects with negative effectiveness ratings. This raises the question why projects whose original vertical logic was altered, and which thus could no longer achieve their intended objectives were continued rather than formally restructured.

- 5.4 Better-quality self-evaluations notwithstanding, there is still room for improvement. The relevance criterion of PCRs requires not only the assessment of alignment of the project with country development needs and Bank strategies, but also with the realities of the country for which these projects were designed. (i.e., economic, environmental, social, political economy, and capacity conditions). Although relevant for the vertical logic of an operation, its implementation and its risk analysis, this information is rather shallow or scarce in PCRs. Furthermore, as in past validation cycles, information on safeguards performance continues to be scarce. On the XSRs side, underachievement of development objectives has been a vexing issue, particularly for FI operations. Explanations for objectives going unmet, when provided, tend to be hypothesis related to external factors and little emphasis is placed on lessons to manage or ways to adjust to external factors.
- 5.5 Neither PCRs nor XSRs are sufficiently focused on enhancing institutional learning. The usefulness of both PCRs and XSRs as a learning tool, and their validations, is hampered by their limited analysis of factors affecting project performance. They commonly provide unclear and inconsistent narratives across evaluative criteria which make it hard to identify the drivers of project results. For example, XSRs' relevance and work-quality challenges were centered around the ill-defined goals and other design flaws; while effectiveness shortcomings were mostly related to unmet targets, rather than to inadequate targets or other results matrix problems. XSRs rarely reflect on these issues.
- 5.6 Lessons drawn from PCRs and XSRs reveal not only the need to strengthen M&E and data collection at the IDBG but also the imperative to document concrete successful actions (to replicate) and unsuccessful actions (to avoid). A large share of lessons drawn focus on how best practices in M&E should have been but were not carried out, and how inadequate data collection hampered the calculation of indicators. The share of lessons that focuses on implementation, aimed at producing knowledge about things we did not know before, is minimal. This reveals a need that operations teams and clients have for strengthened M&E practices and data collection mechanisms. In addition, it exposes the need to document key lessons learned from experience — this is, concrete actions that proved ineffective (and should be avoided) and concrete actions that proved effective (and could be replicated).
- 5.7 Considering these conclusions, OVE recommends:

A. For IDB Management

- 1. Ensure that all future PCRs submitted to OVE integrate alignment with country realities in the relevance assessment.** Most projects with negative relevance ratings were affected by weak vertical logic, which in turn was due to a poor diagnostic assessment or poor alignment with country realities. This then often resulted in implementation problems or cancellation of components, ultimately affecting project results. Therefore, to strengthen the learning component of the PCRs, assessments must consider country realities.
- 2. Ensure that the 2020 PCR guidelines ask for lessons relevant for institutional learning, focused on project elements to replicate or avoid.** To promote institutional learning from SG operations, lessons learned sections of PCRs should include a reflection on successful implementations (e.g., actions that built resilience under unexpected circumstances) and unsuccessful ones. Many lessons stressed the risks of implementation and recommended a more rigorous ex-ante project design to prevent these in future operations. Although a more rigorous project design is undoubtedly necessary, unexpected occurrences are often the rule, not the exception, especially in unstable political and business environments. We seem to know more about how projects should have been designed, but less about what projects did to avoid disappointing outcomes when contexts changed.

B. For IDB Invest Management

- 3. Strengthen the learning component of XSRs, incorporating a reflection on what worked and what went wrong.** According to the XSRs guidelines, the reports are a tool for accountability and learning. However, their learning aspect needs to be reinforced. While low effectiveness ratings reflect unmet objectives, XSRs have focused on analyzing indicators without valuing the need to understand what affects them and how IDB Invest might have responded. The same applies for the lessons-learned section, which mostly focuses on applying already known M&E practices rather than generating knowledge on what worked and what did not.
- 4. Adjust XSR guidelines to ensure that achievements are measured by progress against targets relative to their baseline.** Current guidelines state achievement thresholds based on the absolute proportion of the target not considering baselines. This could lead to irrational results such as an achieved result being rated positive

even if it is below the baseline. OVE recommends that the guidelines be adjusted to reflect the practice that has been initiated during the recently started 2020/2021 XSR exercise which measures progress against targets relative to their baseline.

Office of Evaluation and Oversight - OVE

Established in 1999 as an independent evaluation office, OVE evaluates the performance and development effectiveness of the activities of the Inter-American Development Bank Group (IDB Group). These evaluations seek to strengthen the IDB Group through learning, accountability and transparency.

OVE evaluations are disclosed to the public in accordance with IDB Group policies to share lessons learned with the region and the development community at large.

 iadb.org/evaluation

 facebook.com/idbevaluation

 [@BID_evaluacion](https://twitter.com/BID_evaluacion)