



Project Evaluation

OVE's Review of 2017 Project Completion Reports and Expanded Supervision Reports

Copyright © [2019] 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, 2019**

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

RE-530-2
February 2019

Project Evaluation

OVE´ s Review of 2017 PCRs and XSRs

Office of Evaluation and Oversight



Table of Contents

Acronyms and Abbreviations	vii
Executive Summary.....	viii
Introduction	01
Project Evaluation Framework and Review Process	04
A. Project evaluation framework.....	05
B. Projects reviewed.....	07
C. The project review process.....	09
Public Sector Project Reviews	10
A. Results of validated projects	11
B. Comparison between OVE ratings and Management self-ratings.....	16
C. Quality of PCRs	16
Private Sector Project Reviews	18
A. Results of validated projects.....	19
B. Comparison between OVE ratings and Management self-ratings.....	23
C. Quality of XSRs	25
Conclusions and recommendations	28

Electronic Annexes

Annex I:	<u>List of SG Operations reviewed</u>
Annex II:	<u>OVE ratings and SG Operations</u>
Annex III:	<u>Guidance notes for XSR and PCR quality assessment</u>
Annex IV:	<u>OVE PCR reviews</u>
Annex V:	<u>Comparison of PCR rating results of 2018 validation cycle and those reported in DEO</u>
Annex VI:	<u>List of NSG operations reviewed</u>
Anexo VII:	<u>Overall NSG Performance ratings by subdivisions of validated portfolio</u>

Acknowledgements

This document was prepared by a team composed of Horst Wattenbach (Team Leader), Ana Maria Linares, Jose Claudio Pires, Pablo Alonso, Ulrike Haarsager, Jonathan Rose, Agustina Schijman, Coral Fernandez, Roni Szwedzki, Maria Fernanda Rodrigo Lopez, Odette Maciel, Johanan Rivera, Maria del Mar Carpanelli, Claudia Figueroa, Stephany Maqueda, Nathaniel Russell, Regina Legarreta, Juan Felipe Garcia, Melanie Putic, Xiomara Rojas-Asqui, and the external consultants, Kris Hallberg and Andre Averbug under the overall supervision of Ivory Yong-Protzel (OVE Director).

Acronyms and Abbreviations

CBA	Cost-benefit analysis
CEA	Cost-effectiveness analysis
DEF	Development Effectiveness Framework
DELTA	Development Effectiveness Learning, Tracking, and Assessment
DEM	Development Effectiveness Matrix
DEO	Development Effectiveness Overview
EOM	Early operating maturity (NSG projects)
ERR	Economic rate of return
FI	Financial intermediary
GPS	Good Practice Standards
IDB(G)	Inter-American Development Bank (Group)
IIC	Inter-American Investment Corporation (since 2017 IDB Invest)
MDB	Multilateral development bank
NSG	Non-sovereign-guaranteed
OMJ	Opportunities for the Majority
OVE	Office of Evaluation and Oversight
PBP	Programmatic policy-based operation
PCR	Project Completion Report
PMR	Project Monitoring Report
SCF	Structured and Corporate Finance
SG	Sovereign-guaranteed
SMEs	Small and medium-sized enterprises
SPD	Strategy and Development Effectiveness Department
XSR	Expanded Supervision Report

Executive Summary

Assessing the results of individual development operations is important to learn what worked and what needs to be changed, and to create the basis for strong development effectiveness. Therefore, the multilateral development banks (MDBs) have established systems to measure the effectiveness of their investments. At the Inter-American Development Bank Group (IDBG), the project performance monitoring systems for sovereign-guaranteed (SG) and non-sovereign-guaranteed (NSG) operations consist of three similar building blocks: (i) an *ex-ante* assessment of the project's expected ability to report on results at completion; (ii) project monitoring/supervision reports tracking implementation progress and emerging results, which are prepared twice yearly for SG operations and annually for NSG operations; and (iii) a Management self-evaluation of the results of SG operations at completion (Project Completion Report, or PCR) or of the results of NSG operations when they reach "early operating maturity" (Expanded Supervision Report, or XSR).

The review of project self-evaluations and performance ratings by the independent evaluation office is an integral aspect of the project evaluation system across MDBs. Most MDBs' independent evaluation offices contribute to the project evaluation system by reviewing Management's self-evaluations to validate or, if applicable, override project outcome ratings. MDBs then report the validated ratings emerging from this process in their corporate results reports. The purpose of this process, enshrined in the Good Practice Standards established by the MDBs' Evaluation Cooperation Group, is to ensure the accuracy and credibility of the project performance reporting system. At the IDBG, the Office of Evaluation and Oversight (OVE) has been validating IDB Invest's (formerly IIC's) project self-evaluations since 2001 and reporting results to the Board. At the same time, IDB's self-evaluation system has been developing progressively over the past decade, reinforced by a commitment under IDB's ninth capital increase (IDB-9) that calls on OVE to "validate achieved results in completed projects." This is the second time that OVE is validating the PCRs for operations approved with a Development Effectiveness Matrix (DEM).

The report summarizes the results and findings of OVE's review of XSRs for 43 NSG operations that reached early operating maturity in 2016 and of PCRs for 35 SG operations that closed in 2016. NSG evaluated projects were approved before the new IDB Invest was in place, between 2008 and 2014, while the SG operations were approved between 2008 and 2016. While IDB Invest (and IIC before it) has traditionally completed XSRs for a full cohort of projects that reached early operating maturity in a given year, 2018 was the first year during which IDB also delivered PCRs for essentially all projects that closed in 2016, had been approved with a DEM, and required a PCR. Despite a considerable increase (over 50%) in the numbers of PCRs and XSRs requiring validation in 2018, and intensive consultation between management teams and OVE, the joint effort succeeded in generating the validated data in time for DEO reporting.

Over the last several years, IDBG has adopted an objectives-based system to evaluate the achievements of both SG and NSG projects. The methodology measures project performance against the project development objectives as formulated at approval, assesses how relevant these objectives were, to what extent they were achieved, how efficiently project resources were used, and how sustainable the achieved results are. A performance rating is assigned to each of the four core criteria - relevance, effectiveness, efficiency, and sustainability - and an overall project outcome rating is derived from them. On the SG side, this system has been developed and refined gradually over the last four years, with the adoption and revision of three consecutive sets of PCR guidelines between 2014-2018. On the NSG side, in 2015 the IDBG introduced and tested new guidelines for the preparation of XSRs using the objectives-based methodology, and an update of the guidelines was agreed between IDB Invest and OVE in 2017. These new guidelines introduced, for example, interim rules for those NSG FI operations which had been approved without appropriate approvals objectives.

Applying the methods embraced by the most recent XSR and PCR guidelines, OVE determined that 16 of 35 SG operations and 25 of 43 NSG operations achieved an overall outcome rating of Partly Successful or higher. All but one of the SG operations and 32 of the 43 NSG operations scored Satisfactory or higher on relevance, indicating that they were well aligned with country development challenges and that their design was internally consistent. Projects scored less well on effectiveness and efficiency. Only six of 35 SG operations and 21 of the 43 NSG operations presented evidence of reaching the majority of their stated objectives, thus scoring a Satisfactory or higher rating on effectiveness. The lower SG project ratings are due to a combination of failure to achieve stated development objectives and lack of information on some results indicators. As the 2018 PCR guidelines take effect, the latter factor is expected to diminish, provided project teams are made

aware of the need to ensure that information contained in the original results matrix (or the matrix adopted within 60 days after eligibility) is regularly collected and reported on in PCRs. Twenty of the 35 SG operations and about two-thirds of the NSG operations were rated Satisfactory or higher on sustainability, suggesting that continuation of project benefits is at considerable risk for the remaining projects.

NSG operations were also assessed on three non-core criteria: additionality, investment outcome, and IDBG work quality. The majority of NSG projects performed well on additionality and investment outcome: among the 43 validated projects, 34 were rated Satisfactory or higher on additionality and 38 on investment outcome. The investment outcome rating suggests that the IDBG did well in selecting and structuring operations that contributed positively to its financial success, and that most loans get paid as scheduled. However, the quality of IDBG's work for the reviewed operations was uneven. Only one-third of the NSG operations demonstrated high work quality during structuring and appraisal as well as during supervision. Shortcomings were particularly marked during structuring and appraisal. The latest guidelines for the preparation of PCRs also provide that Bank and borrower performance be reviewed as non-core criteria. However, as PCRs reviewed during the 2018 validation exercise did not yet contain information on these aspects, they were not included in this year's validations.

A key purpose of the independent evaluation office's validation of project outcome ratings is to ensure the credibility of the self-evaluation system. For this reason it is useful to compare OVE's final project outcome ratings with Management's own ratings. On the NSG side, there was noticeable divergence between the project outcome ratings in XSRs prepared by IDB Invest and the final ratings assigned by OVE. OVE ratings coincided with Management's self-ratings for only 17 of 43 projects; OVE downgraded 23 projects and upgraded three. Following further discussion and review of evidence, IDB Invest indicated agreement with all but three of OVE's final outcome ratings. IDB invest teams put significant effort into applying the jointly agreed upon guidelines for FI operation which helped limit ratings discrepancies between OVE and management for such operations.

A similar comparison between OVE ratings and IDB's self-ratings is not yet possible for SG operations. PCRs were prepared using various guidelines and methods, and, as stipulated in the 2018 PCR guidelines, OVE applied the latter in its validations; thus strict comparability is not possible. Starting with 2019, IDB and OVE ratings are expected to be comparable, as PCRs will be prepared using the 2018 guidelines.

Over the past year, significant progress has been made in consolidating the IDBG's project self-evaluation system. The PCR and XSR guidelines have become a satisfactory basis for preparation of PCRs and XSRs. This validation exercise has, nevertheless, pointed to several areas in which further clarification, best documented in agreed guidance notes and training of both IDBG staff and OVE evaluators would be beneficial to ensure consistency in evaluation and validation. Furthermore, as the number of self-evaluations is expected to increase considerably over the coming years, particularly on the IDB side, the delivery timetable and validation process will need to be further refined between Management and OVE to ensure that OVE can feasibly validate all self-evaluations in time to deliver final ratings as regular inputs to the Development Effectiveness Overview.

In light of the experience of this year's validations, OVE recommends the following:

For IDB Invest Management:

- (i) In collaboration with OVE, develop guidance notes and train XSR authors and OVE evaluators in improved financial and economic analysis methods, to ensure common standards in terms of clearly structured calculations, homogenous methodology, and transparent assumptions.
- (ii) Agree with OVE on areas in which additional guidance notes to support the practical application of XSR guidelines would be beneficial and develop such notes in collaboration with OVE.
- (iii) Agree with OVE on a process of independent review and verification of the project pipeline to determine the list of operations reaching early operating maturity, and subject to XSR preparation and validation every year.
- (iv) Agree with OVE on a delivery schedule of final XSRs that ensures that XSRs submitted for validation have passed all internal quality controls, taking into account the increased XSR numbers and deadlines for corporate reporting.

For IDB Management:

- (i) In collaboration with OVE, train project team leaders and PCR authors on the 2018 PCR guidelines to ensure that teams are aware of the requirements to monitor performance and report in accordance with the originally approved results matrix (adjusted within 60 days of eligibility, if applicable) and that all new PCRs are prepared according to 2018 guidelines. Make self-ratings an integral part of PCRs.

- (ii) Develop complementary sector-specific guidance notes on the preparation of cost-benefit analysis and cost-effectiveness analysis as an integral part of PCR preparation.
- (iii) Adopt interim guidelines for evaluating financial intermediary operations consistent with those developed for NSG projects, to bridge the period during which FI operations approved with beneficiary lists rather than sub-portfolio-specific objectives are being implemented.
- (iv) Agree with OVE on a delivery schedule of PCRs in view of increased PCR numbers and deadlines for corporate reporting.



01

Introduction

- 1.1 Assessing the results of individual development operations is important to learn what worked well and what needs to be changed, and to create the basis to for strong development effectiveness. It also helps promote accountability for results. Therefore, the multilateral development banks (MDBs) have established systems to measure the effectiveness of their investments. In 2008, the Inter-American Development Bank (IDB) launched its Development Effectiveness Framework (DEF) to generate a body of knowledge about what works in meeting the region’s development challenges. The IDB’s ninth capital increase (IDB-9)¹ reinforced the commitment to enhanced reporting on the development effectiveness of Bank operations, when the institution pledged to “improv[e] the results measurement framework to identify for every project whether specific and tangible results have been achieved” (CS-3868-1). At the same time the Bank established the target that 80% of projects should achieve satisfactory development results at completion by the year 2019, and it reports on annual progress toward this target in its Development Effectiveness Overview (DEO). Until the IDBG’s recent merge-out, the DEF applied to sovereign-guaranteed (SG) operations as well as operations approved by the former Structured and Corporate Finance department (SCF) and the former Opportunities for the Majority initiative (OMJ).
- 1.2 For IDB Invest and its predecessor the Inter-American Investment Corporation (IIC), the more systematic tracking of development results dates back to 2001 (CII/RE-1). The IIC established a project evaluation system and function in 1999 (CII/GN-141, CII/GN-141-2), contracted the Office of Evaluation and Oversight (OVE) to support this effort in 2000, and in 2001 started to self-evaluate its operations following the Good Practice Standards for Private Sector Operations issued by the MDBs’ Evaluation Cooperation Group. Since the 2016 merge-out of the IDBG’s private sector operations into IDB Invest, the Development Effectiveness Learning, Tracking, and Assessment tool (DELTA) (CII/PP-169, CII/PP-180) has replaced the two systems that were previously used in SCF/OMJ (Development Effectiveness Matrix, or DEM) and IIC (Development Impact Assessment Scorecard). Thus, DELTA has become the common system for ex-ante scoring of development results and additionality of operations within the broader framework of the IIC’s “Renewed Vision” for private sector operations (CII/GN-2807-2).

¹ Report on the Ninth General Increase in the Resources of the Inter-American Development Bank, AB-2764.

- 1.3 The project performance monitoring and evaluation systems for SG and non-sovereign-guaranteed (NSG) operations consist of three similar building blocks: (i) an ex-ante assessment of the project’s expected ability to report on results at completion (DEM for SG operations, DELTA for NSG operations); (ii) project monitoring/supervision reports tracking implementation progress and emerging results (Project Monitoring Reports prepared twice yearly for SG operations and Project Supervision Reports prepared annually for NSG operations); and (iii) a Management self-evaluation of the results of SG operations at completion (Project Completion Report, or PCR) or of the results of NSG operations when they reach “early operating maturity” (Expanded Supervision Report, or XSR).²
- 1.4 The review of project self-evaluations and performance ratings by the independent evaluation office is an integral aspect of the project evaluation system across MDBs. Most MDBs’ independent evaluation offices contribute to the project evaluation system by reviewing Management’s self-evaluations to validate or, if applicable, override project outcome ratings. MDBs then report the validated ratings emerging from this process in their corporate results reports. The purpose of this validation, enshrined in the Good Practice Standards (GPS) established by the MDBs’ Evaluation Cooperation Group for private and public project evaluations, is to ensure the accuracy and credibility of the project performance reporting system. At the IDBG, OVE has been validating IDB Invest’s (formerly IIC’s) project self-evaluations since 2001 and reporting results to the Board. At the same time, IDB’s self-evaluation has been developing progressively over the past decade, reinforced by a commitment under IDB’s ninth capital increase (IDB-9) that calls on OVE to “validate achieved results in completed projects.” This report represents only the second time that OVE is validating PCRs for operations approved with a DEM.³
- 1.5 This report summarizes the results and findings of OVE’s review of the XSRs completed by IDB Invest for NSG operations that reached early operating maturity (EOM) in 2016 and of the PCRs that IDB completed for projects that closed in 2016. It also points out areas in which the project evaluation systems need further refinement to ensure that they can help generate evidence-based results.

2 The exact definition of “early operating maturity” varies by project type, but it 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 2 of the IDBG XSR guidelines for further details.

3 In 2015, OVE reviewed 12 PCRs only prepared on a pilot basis (see RE-488). The first full validation of PCRs prepared by the end of 2016 was completed by OVE in 2017 (see RE-250).



02

Project Evaluation Framework and Review Process

A. Project evaluation framework

- 2.1 IDBG is in the forefront among major MDBs, in that it has adopted a unified, objectives-based system for evaluating and rating both public and private sector operations. On IDB's side, the system to assess and rate project performance against stated project development objectives has been developed and refined gradually over the last four years, and three consecutive sets of PCR guidelines have been adopted and revised between 2014 and 2018. Each set built on an assessment by Management and OVE of strengths and weaknesses during the preparation and validation of a set of PCRs that applied them. The most recent guidelines, adopted by IDB's Operations Committee in January 2018, incorporate suggestions made in OVE's last validation report (RE-250). This last set of guidelines⁴ stipulates that OVE will start applying the methodology set out in them immediately, regardless of whether PCRs were prepared under earlier guidelines. Therefore, the set of PCRs that OVE validated in 2018, which this report covers, is based on these new guidelines, while the underlying PCRs were prepared using either the 2014 or the 2016 guidelines. OVE's application of the 2018 guidelines has the benefit of putting in practice the ratings guidance provided in the new guidelines and measuring with the same yardstick the outcomes of all projects that closed in 2016 (and had been approved with a DEM).⁵ Furthermore, the 2018 PCR guidelines are largely aligned with IDB Invest's XSR preparation guidelines, allowing for differences in the nature of SG and NSG operations (see paragraph 2.4).
- 2.2 In 2015 the IDBG introduced and tested new guidelines for the preparation of XSRs for NSG operations. These guidelines established the use of the objectives-based methodology to evaluate the performance of NSG operations. In view of the recommendations issued by OVE in RE-522, an update of the guidelines was agreed between the IDB Invest Strategy and Development Department and OVE. This update further aligns the assessment criteria and ratings methodology. IDB Invest applied the revised guidelines in the preparation of the XSRs that were subject to OVE validation in 2018, and OVE applied them in its validation of the 43 XSRs summarized in the present report.
- 2.3 The objectives-based evaluation methodology takes as the starting point the stated development objectives that an operation aimed to achieve. Thus, rather than evaluating project

4 GN-2489-10 of 30 May 2018 presented the final version of the 2018 PCR guidelines as Annex.

5 Of the 35 PCRs submitted to OVE for validation, 15 were prepared following the 2014 guidelines and 20 used the 2016 guidelines.

results in the abstract, the objectives-based methodology measures project performance against the project development objectives that were formulated at approval, assessing how relevant these objectives were, to what extent they were achieved, how efficiently project resources were used to achieve them, and how sustainable the achieved results are. A performance rating is assigned to each of these four core criteria, and an overall project outcome rating is derived from them (Table 2.1).

Table 2.1. Core project performance assessment criteria

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 economic benefits exceed costs of capital; project effects on key economic stakeholders
<ul style="list-style-type: none"> • Safeguards performance • Assessment of environmental, social, technical, financial, economic, and political risks to continuation of project development results 	Sustainability	<ul style="list-style-type: none"> • Safeguards performance • Assessment of unmitigated risks to continuation of project results

Overall Project Outcome

Source: OVE.

2.4 The development outcome rating systems for SG and NSG projects is now based on the same structure and aggregation rules. There are four core criteria: relevance, effectiveness,⁶ efficiency, and sustainability. Core criteria are rated on a four-point scale ranging from Excellent to Unsatisfactory. The core ratings are aggregated in a weighted average to derive an overall project outcome rating, with a weight of 20% each for relevance, efficiency, and sustainability and 40% for effectiveness. The resulting overall outcome is rated on a six-point scale ranging from Highly Successful to Highly Unsuccessful. While the overall

⁶ The only exception to this rule is SG programmatic policy-based operations, for which efficiency is not assessed. Their overall outcome rating is based on relevance (20%), effectiveness (60%), and sustainability (20%).

framework to arrive at a project outcome rating for SG and NSG operations is thus similar, its application differs somewhat, because of the different nature of SG and NSG operations.

- 2.5 In addition to the four common core criteria, XSR and PCR guidelines stipulate a set of non-core criteria that project self-evaluations are to cover. On IDB's side, the 2018 PCR guidelines introduced the assessment of Bank performance and borrower performance as non-core criteria. The first is an assessment of the quality of services provided by the Bank during the life of a project. The second focuses on the effectiveness with which the borrower discharged its responsibilities. This further aligns IDB's PCR guidelines with the GPS and adds to compliance with the DEF, which stipulates that Bank performance during the life of the project will be assessed.⁷ Since the PCRs submitted to OVE in early 2018 were drafted according to pre-2018 PCR guidelines, they did not yet contain these sections. Hence OVE could not validate these two aspects during the current exercise. This is expected to change with the next crop of PCRs that OVE will validate in 2018/19. The non-core criteria of NSG operations have remained unchanged in the 2018 XSR guidelines and include IDB Invest's financial and non-financial additionality, investment outcome, and work quality (with sub-ratings in each area for screening/appraisal/structuring and supervision), all of which are rated on a four-point rating scale.

B. Projects reviewed

- 2.6 On the private sector side, OVE validated the XSRs for 43 NSG operations that reached early operating maturity in 2016. The performance of these operations was presented in 41 XSR reports as two reports covered two loans each. These operations were approved between 2008 and 2014, i.e. before the new IDB Invest was in place: 25 had been approved by the former IIC, 14 by IDB's SCF, and four by OMJ; and 27 were financial intermediary (FI) operations, five were for infrastructure, and 11 for corporate investments (Table 2.2). The reviewed NSG portfolio amounted to US\$934 million in approvals, of which US\$659 million (70.5%) went to FI projects, US\$136 million (14.5%) to corporate operations, and US\$139 million (15%) to infrastructure. The validations covered 15 IDB member countries (some with multiple operations) and six regional operations (see Annex VI for the project list).

⁷ IDB Results Framework 2012-2015, paragraph 6.6.

Table 2.2. Distribution of NSG Operations reviewed

Specific Sector	Number of Projects				Loan Approvals (US\$ million)
	Private Sector Windows			Total number	
	Former IIC	OMJ	SCF		
Corporates	10	0	1	11	135.8
Infrastructure	3	0	2	5	139.2
Financial Institutions	12	4	11	27	659.3
Total	25	4	14	43	934.2

Source: OVE.

2.7 On the public sector side, PCRs were for the first time presented for a coherent cohort of projects and programs that closed in 2016. Previously, the PCRs delivered to OVE and hence its validations covered various project closure years. In 2018, the Strategy and Development Effectiveness Department (SPD) ensured that PCR delivery focused on projects that closed in 2016 for which OVE validation was required (i.e., PCRs prepared according to the 2014 or later guidelines for operations approved with a DEM, subject to certain exclusions)⁸. Of the 41 PCRs completed for operations that closed in 2016, six had already been validated and reported on in OVE's 2017 validation report, so that this report covers 35 PCRs for operations closed in 2016. The 35 validated PCRs covered 49 loans in 20 countries: 26 investment operations (including one grant operation) and 23 policy-based loans supporting 9 policy-based programs (PBPs). Since PBPs are evaluated and rated at the program level, the remainder of the report will count each PBP as one single unit, regardless of how many loans it covered.

2.8 The 35 PCRs cover operations from all four sector departments, with 10 operations each falling into the Infrastructure and Energy, Social Sector, and Institutions for Development departments and five into the Climate Change and Sustainable Development Department (Table 2.3). By sector, the largest number of PCRs were for operations in the Water and Sanitation sector (six), followed by Social Investment and Modernization of the State (four each), Education, Health, Urban Development and Housing, and Financial Markets (three each) (see Annex I for a list of the PCRs validated). The total approved lending covered by the 2018

⁸ According to information provided by SPD, 84 operations closed in 2016, 65 of which had been approved with a DEM. Of these, 15 will require a PCR only upon completion of later elements of a series or supplemental financing and nine were cancelled before reaching disbursement levels requiring a PCR. Since this 2018 validation cycle closed, two additional projects were identified by Management that closed in 2016, illustrating that final reporting by closure year is a dynamic process.

PCR validations amounted to US\$3,445 million. Because of the limited number of PCRs validated in each sector thus far, it is not yet possible to undertake performance analysis across sectors.

Table 2.3. Sector department distribution of PCRs reviewed

Sector	SG Operations (# of projects)	Amount approved (US\$ millions)
Infrastructure and Energy (included Water & Sanitation)	10	535.6
Social sector (Education, Health, Gender and Labor Markets)	10	1,347.8
Environment, Rural & Urban Development and Housing	5	208.8
Reform, Financial Markets and Trade & Competitiveness	10	1,352.7
Total	35	3,444.9

Note: The nine PBPs included in the validations count as one PCR each in this table, but cover 23 loan approvals.

Source: OVE.

C. The project review process

2.9 In general terms, the same validation process applied to private and public sector projects. An OVE evaluator was assigned to review each PCR and XSR and prepare an assessment report. A small group of peer reviewers was assigned to review batches of validations to ensure consistency across project assessments. For a part of the NSG operations—the corporate sector XSRs—OVE had to seek additional clarification from IDB Invest to be able to review and validate the submitted calculations on financial and economic efficiency. When available, IDB Invest counterparts readily provided the additional information and/or clarifications which allowed OVE to complete the draft validations. OVE’s draft assessment reports with project-specific ratings were shared with IDB and IDB Invest Management for comments. For NSG operations, written comments were received on 22 XSR validations (54%). The share of projects receiving written comments was lower than in the previous year, possibly an indication of less divergence between self-evaluations and OVE validations, as the preparation of XSRs using the new guidelines takes hold. On the SG side, the validation process triggered substantially more interaction with project teams than last year, with OVE receiving requests for meetings to discuss validations for 19 of the 35 PCRs and written follow-up comments for 23 validations. OVE then finalized its validations, taking into account Management comments and any additional evidence provided.



03

Public Sector Project Reviews

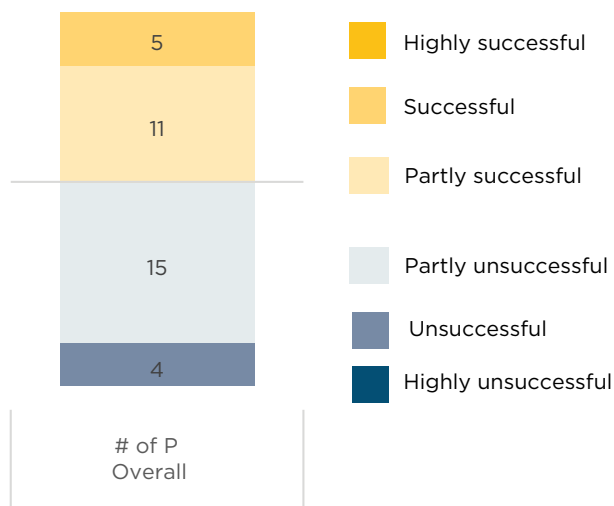
A. Results of validated projects

1. Overall outcomes

- 3.1 The project outcome rating provides a snapshot of an operation's overall performance. The outcome rating summarizes how relevant an operation and its objectives were at approval, to what extent it has achieved its development objectives, how efficiently the project converted resources into results, and how likely it is that the achieved results will be sustained. Outcome is rated on a six-point scale ranging from "Highly Successful" to "Highly Unsuccessful". As section IIA explained, the overall rating is a weighted average of the ratings of the four core criteria (the validations of PCRs are presented in Annex IV).
- 3.2 Of the 35 SG projects and programs, 16 were rated Partly Successful or higher on overall development outcome (Figure 3.1). Five of the nine PBPs (56%) and 11 of the 26 investment operations (46%) were rated Partly Successful or higher. However, the number of operations in each category is too small to draw any meaningful comparative conclusions across lending instruments. Unsuccessful ratings were mostly driven by failure to demonstrably achieve development objectives (i.e., low effectiveness ratings), as discussed further below. None of the projects fell into the rating categories of "Highly Successful" or "Highly Unsuccessful".

Figure 3.1
Overall project
outcome ratings
(N=35)

Source: OVE.



2. Core rating criteria

- 3.3 Across the core criteria that contribute to the overall outcome rating, the reviewed projects scored highest on relevance and lowest on effectiveness. As discussed in paragraph 2.4, the core

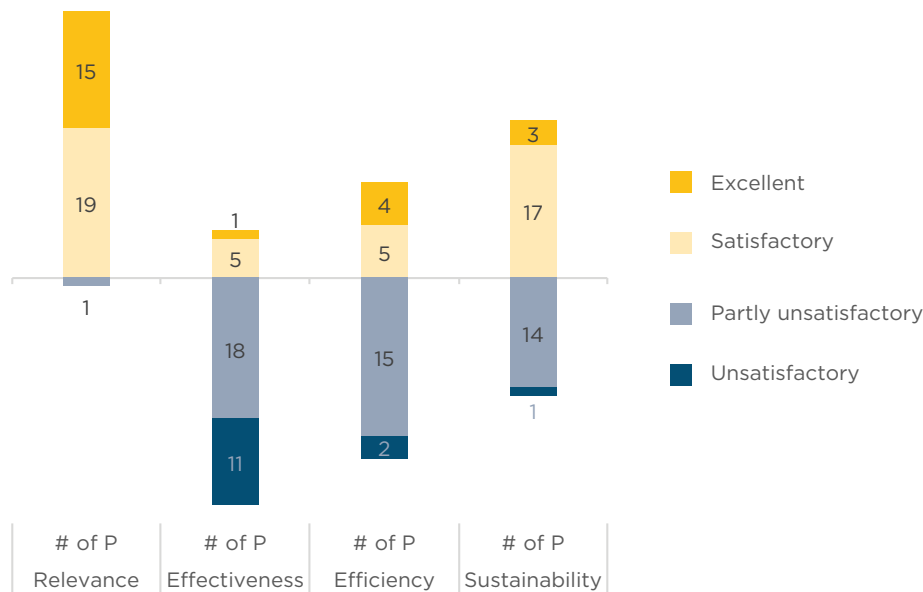
criteria of relevance, effectiveness, efficiency, and sustainability are rated on a four-point rating scale (Figure 3.2). Only one operation was rated less than Satisfactory on relevance, but 29 were rated less than "Satisfactory" on effectiveness, 17 on efficiency, and 15 on sustainability.

Figure 3.2

Project ratings by core criteria (N=35)

Source: OVE.

NOTE: Nine PBL operations are not rated for efficiency.



Relevance

3.4 The high share of projects with "Excellent" or "Satisfactory" relevance ratings indicates that the reviewed operations were well aligned with the country development challenges and realities and that their design was internally consistent. The one operation that was rated less than "Satisfactory" on relevance had weaknesses in showing the evidence of the underlying development problem, and of how the loan would fill the financing gap for target groups. Relevance shortcomings generally related to the internal consistency of the project logic and its matching with the core development problem.

Effectiveness

3.5 Effectiveness measures the extent to which an operation has achieved its stated development objectives. To assess effectiveness, OVE started the validation with a review of the operation's development objectives as stated in the loan document or loan contract and then saw how the outcome indicator(s) in the results matrix aligned with each of these objectives. Based on this mapping, the achievement ratio of each outcome indicator is assessed against targets in the results matrix defined at approval or within 60 days after the operation reached eligibility. Indicator achievement ratios are aggregated to determine the achievement

ratio for each objective, which then translates into a rating for each objective. Individual objectives' ratings are then aggregated to derive an overall effectiveness rating.

- 3.6 Of the 35 operations reviewed, six achieved an effectiveness rating of "Satisfactory" or higher. Only one operation demonstrably achieved all its stated development objectives and was rated "Excellent" on effectiveness, and five more showed that they achieved more than half their stated objectives and were thus rated "Satisfactory". The remaining 29 operations failed to demonstrably achieve the majority of their objectives. Low effectiveness ratings can be caused by a failure to achieve the target indicators, or by lack of information on these indicators.⁹ Of the 29 operations with less than "Satisfactory" effectiveness ratings, clear shortcomings in achieving stated objectives was an important reason in 25 (86%) and lack of information on achievements against the valid indicators was an important reason in 16 (55%).¹⁰
- 3.7 The effectiveness ratings for PBPs warrant additional discussion. PBPs are approved as a series of loans over time, and the results matrix for the program as a whole with end-of-program targets may be changed with each new loan approved in the series. In all nine PBPs reviewed during this validation exercise, the results matrix was adjusted to some degree between the first and the last operations of the program, with the end-of-program targets in the last results matrix often being set less than a year before program closure.¹¹ Revising targets downward increases the chances of reaching the target values. The effect is visible in operations that rate progressively higher against the matrix approved with each new loan in a series. OVE analyzed the extent to which changes to the results matrices across the various loans in a series affected the achievements at the end of the series: measured against the end-of-program targets of the first results matrix in each series, only one PBP would have been rated Satisfactory, while measured against the end-of-program targets of the last results matrix in each PBP series, five would

9 A lack of available information in turn is also caused by several factors. Some projects did not collect the information required to report on some of the agreed indicators. Others adjusted the indicators very late during implementation and therefore provided information on indicators and target values established long after the deadlines established in the guidelines for such modifications.

10 Among the 25 operations with shortcomings in achievement of objectives and the 16 with lack of information on indicators, both factors simultaneously contributed to low effectiveness ratings for 12 operations.

11 Matrix changes include different combinations of adding/removing indicators for the same objectives during a series, but also adjustments in the indicator values for the end-of-series moment with each new loan approval, which is equivalent to shifting the goalpost as the target moment gets closer. The last results matrix for seven of nine PBPs was approved less than one year before program closure.

have been rated high (three "Excellent" and two "Satisfactory").¹² To account for possible downward adjustments as the series progresses, the 2018 PCR guidelines stipulate that if the results matrix changes over the course of the program, the program will be evaluated against the end-of-program targets in each results matrix and then an average effectiveness rating will be derived. Applying this method resulted in two of the nine PBPs being rated Satisfactory on effectiveness, six "Partly Unsatisfactory", and one "Unsatisfactory".

- 3.8 The validations also illustrate the importance of setting reasonably ambitious outcome targets to avoid illogical results at completion. A water supply project aimed (among other objectives) to reduce extremely high losses of drinking water by replacing a large share of the water network, using a technical assessment that was to determine priority intervention locations. The project failed to carry out this assessment in time, and only 49 km instead of 129 km of main water pipes were replaced. Therefore, at closure, the project clearly failed on the output delivery of "replacement of water mains" but had defined the outcome target ("reduction of water losses") so low that the target improvement endorsed at approval was still fully reached. The case illustrates the importance of defining targets at approval with reasonable, but also sufficient, ambition.

Efficiency

- 3.9 The rating on efficiency is based on the expectation that projects should carry out either a cost-benefit analysis (CBA) or a cost-effectiveness analysis (CEA) at closure. If a CBA is carried out at completion, the assigned rating is based on the CBA results compared to the opportunity cost of capital. Alternatively, a comparison with the cost-effectiveness of alternative approaches to achieving project results may be carried out. According to the PCR guidelines, a project cannot achieve an efficiency rating of Satisfactory or higher if the PCR does not present a CBA or CEA. The project efficiency rating is then derived from certain indicators in the Project Monitoring Reports (PMRs).¹³ Since PBPs (nine PCRs) do not require an efficiency analysis, the following discussion on efficiency ratings refers to the 26 PCRs for investment operations only.

¹² The ratings of all PBPs were lower using the first matrix, except one that would have been rated Unsatisfactory in any event.

¹³ In the absence of CBA or CEA, the efficiency rating is derived from the ratings of the CPI and SPI indexes in the last PMR, with projects with a "Satisfactory" PMR rating in relation to these indexes achieving a "Partly Unsatisfactory" efficiency rating and those with an alert or problem PMR rating achieving an "Unsatisfactory" efficiency rating.

3.10 Nine of the 26 investment operations reviewed achieved an efficiency rating of Satisfactory or higher. Six PCRs were rated based on only their CBA analysis, and two based on their CEA. Eight PCRs had performed either a CBA or a CEA for only part of investment costs so that the validation had to go beyond the CBA/CEA results to look at other determinants of efficiency for the entire project. Some projects with a CBA were downgraded because of methodological shortcomings in calculations which made the seemingly high economic rates of return (ERRs) presented in the PCR questionable.¹⁴ Downgrades were also necessary when the validity of the presented results could not be confirmed or when calculations were based on highly abstract models without providing actual evidence on project benefits.¹⁵ Such cases point to a need for better guidance on the minimum required evidence of actual project benefits to complement academic models for estimating ERRs.

Sustainability

3.11 In about two-fifths of the projects reviewed there is a risk that project benefits will not be sustained. The sustainability rating considers risks to the continuation of project-generated benefits, including any unmitigated social and environmental risks arising from underperformance on environmental and social safeguards. Of the 35 operations, 20 were rated Satisfactory or higher on sustainability, suggesting that continuation of project benefits is at risk for the other 15 projects. In 13 of these 15 projects, lack of financial sustainability was a key aspect threatening the continuation of results. Information on safeguards performance was scant in essentially all PCRs, a reflection of the lack of systematic monitoring and reporting on safeguards performance and results throughout project implementation Bank-wide.¹⁶ Of

14 A road rehabilitation project presented an ERR calculation with results sufficiently high to merit an efficiency rating of "Excellent". However, a review of the calculations indicated serious methodological shortcomings in the calculation, so that OVE corrected for the mistakes and the results dropped to "Partly Unsatisfactory". An adjustment in the opposite direction was carried out in the case of a rural water supply project, which had ERR results at the margin between "Partly Unsatisfactory" and "Satisfactory". The PCR had not argued explicitly for specific standards to be exempted from the general benchmark (opportunity cost of capital). However, the project in question had explicitly targeted dispersed settlements in rural areas, which could quite clearly not expect an ERR performance in line with the standard opportunity costs of capital. The example illustrates the need to establish clear guidance notes on which sectors do require ERR calculations as well as criteria when circumstances warrant adjustments from the standard ERR benchmarks.

15 ERR calculations were presented for a project that rehabilitated and expanded water supply services in rural areas. The project documents indicated that the situation before the project was that beneficiaries had to resort to water provision from tanker truck services and therefore had real opportunity costs with which the end-of-project situation could be compared. However, the *ex post* ERR calculation replicated an *ex-ante* model, in assuming water demand elasticities from neighboring countries and using assumed willingness to pay for water provision, but it did not present information on real water demand by beneficiaries or consider the savings realized by the beneficiary population from avoided expenditure for trucked water.

16 See OVE's recent evaluation on environmental and social safeguards, RE-521-1.

the 20 PCRs for projects classified as being in environmental risk category A, B, or FI, only three reported on the effects of some environmental and social shortfalls or results, while five others stated that the environmental and social management or local environmental and social requirements had been complied with, without providing evidence.

B. Comparison between OVE ratings and Management self-ratings

3.12 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 PCR ratings. Such a comparison is, however, difficult for the 35 projects subject to this year's OVE validation, as PCRs were prepared under two sets of older guidelines. To attempt such a comparison in the context of the Development Effectiveness Overview (DEO), SPD developed conversion factors for project ratings that had been prepared using the 2014 guidelines. However, beyond the conversion from one rating scale to another, there are other methodological differences underlying the project ratings that considerably limit the comparability of (i) Management's ratings for the 15 projects for which PCRs were prepared using the 2014 guidelines; (ii) Management's ratings for the 20 projects for which PCRs were prepared according to the 2016 guidelines; and (iii) the ratings for both sets of projects derived by OVE applying the 2018 PCR guidelines. Going forward, as the PCRs are prepared using the 2018 guidelines, a comparison of Management's ratings with OVE's ratings will become more meaningful.

C. Quality of PCRs

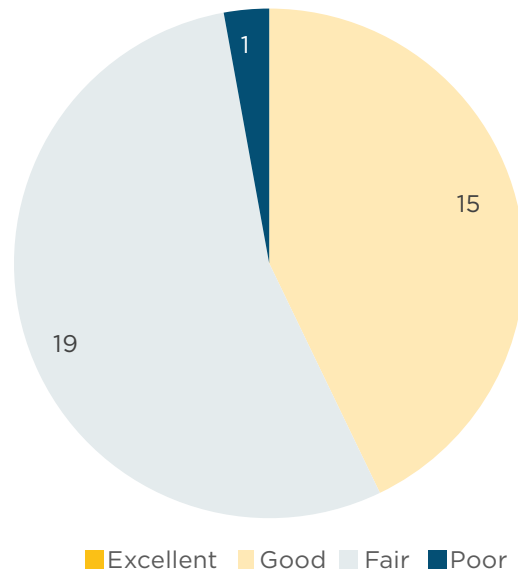
3.13 The quality of PCRs was variable. OVE assesses the quality of the PCRs on a four-point rating scale ranging from "Excellent" to "Poor" (see Annex III). Fifteen reports were rated "Good", 19 were rated "Fair" – that is, they were overall of reasonable quality but had some shortcomings or one key shortcoming – and one was rated "Poor", meaning it had several major shortcomings (Figure 3.3). Most PCRs were candid and did well in analyzing the vertical logic of projects and in discussing implementation challenges. A frequent shortcoming among the PCRs that were rated less than "Satisfactory" was that they did not clearly analyze project achievements around a project's stated development objectives, but instead either discussed the achievements of project components or contained a general discussion of project

achievements. To some extent, this shortcoming may be linked to the PCRs' having been prepared using older guidelines. Going forward, as the new approach to preparing PCRs takes hold, it will be important to ensure that project teams clearly understand that the evaluation of effectiveness needs to be centered around the achievement of individual objectives.

Figure 3.3

Rating of PCR quality

Source: OVE.





04

Private Sector Project Reviews

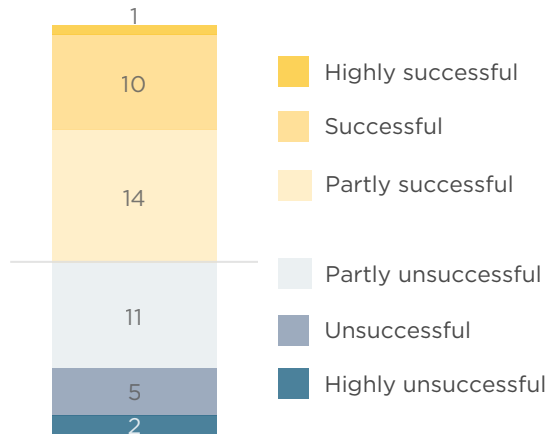
A. Results of validated projects

1. Overall outcomes

4.1 Of the 43 NSG operations reviewed, 25 achieved outcome ratings of "Partly Successful" or higher (Figure 4.1). Financial intermediary (FI) operations constituted the largest group (27 projects) among the 43 evaluated operations and accounted for 12 of the 18 operations rated "Partly Unsuccessful" or lower. A deeper analysis of rating differences by project categories (e.g., by which private sector approved it or by sector – FI, corporate, infrastructure) has limitations given the small sample size. Over time, as more validations are completed using the objectives-based methodology, it will become possible to undertake comparisons across major types of NSG operations and sectors by pooling validations across several EOM years (see Annex VII).

Figure 4.1
Overall project outcome ratings (N=43)

Source: OVE.

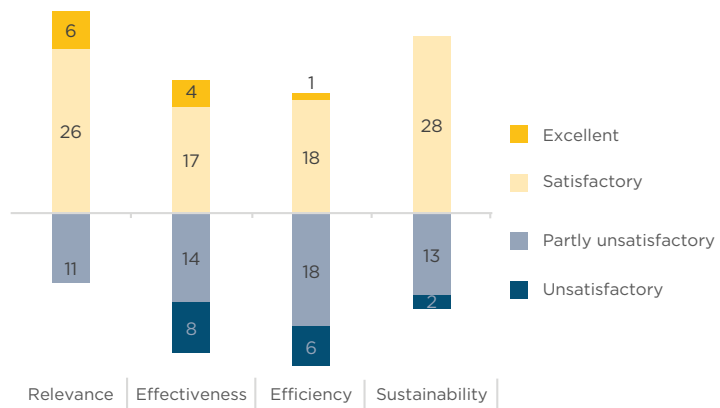


2. Core criteria

4.2 Across the four core criteria contributing to the overall outcome rating, projects scored best on relevance, followed by sustainability, effectiveness, and efficiency. Of the 43 projects reviewed, 32 were rated Satisfactory or higher on relevance, 21 on effectiveness, 19 on efficiency, and 28 on sustainability (Figure 4.2).

Figure 4.2
Project ratings by core criteria (N=43)

Source: OVE.



Relevance

- 4.3 Almost three-quarters of NSG projects were rated "Satisfactory" or higher on relevance—that is, they aimed to achieve development objectives that were relevant to the country's development challenges, were aligned with the approving private sector windows' business priorities, and did not have any substantial shortcomings in project design. Of 11 projects rated "Partly Unsatisfactory" on relevance, eight were FI operations (30% of 27), a higher share than for projects with corporate clients (three of 16, 19%). The relevance of FI operations was in part affected by the existence of alternative channels of subsidized loans to the target groups in countries with highly regulated interest rates. One project rated low on relevance aimed at providing commercial renewable energy packages for poorer families, but the relevance of the project was hampered by, among other things, the lack of a market demand study to inform project design. In another project, relevance was affected by a lack of clarity over who the target beneficiaries would be—whether the target group was small and medium-sized enterprises (SMEs) that in turn should support "base of the pyramid" producers, or whether the SMEs themselves were defined as the target group.

Effectiveness

- 4.4 About half of the NSG operations (21 of 43) achieved most of their development objectives and were thus rated "Satisfactory" or higher on effectiveness. The share of operations achieving a "Satisfactory" or higher rating on effectiveness was similar for FI operations and the rest of the portfolio.
- 4.5 The common aim of the 27 FI operations was to improve access to finance for micro, small, and medium enterprises. Some targeted more specific groups, such as female-headed enterprises, and some also aimed to expand access to mortgage financing for middle- and lower-middle-income households. Projects with more ambitious targets faced more difficulties in achieving a high effectiveness rating. This also applied to the degree with which a project was expected to provide access to financing for difficult-to-reach target groups (women-led SMEs, agricultural SMEs, and environmentally sustainable investments), rather than merely expanding access by standard SMEs. For FI operations, a common earlier practice had been to have objectives like "the project will finance X SMEs with amounts of US\$Y million." Given the fungibility of money, in most cases it is not meaningful to measure achievement of this objective by a list (selected by the FI) of enterprises reportedly funded by IDB IIC, but 16 (of 27) FI operations did. Following

the agreed recommendations of OVE's FI evaluation (RE-486-2), project evaluation guidelines were adapted to include interim guidelines for FIs that focus instead on the extent to which the FI's relevant portfolio increases (in comparison to the size of IDBG's loan). Using the revised guidelines, about half of the FI projects (13 of 27) were rated Satisfactory or higher on effectiveness. In most projects (22 of 27), IDB Invest was able to obtain information about the relevant portfolio, including in 12 whose objectives had been based on a list of projects.

- 4.6 Some projects included objectives without clear metrics, making the effectiveness assessment difficult and leading to rating downgrades. For example, two projects included a "demonstration effect" that would motivate other FIs to follow their path. But as no indicators were included, progress on this matter was difficult to assess.¹⁷ Lack of information on at least one objective was a concern for 15 of the 27 FI operations, with 11 rated low on effectiveness. A review of FI operations found 15 projects that clearly failed to reach any objectives, and nine projects with a lack of information on at least one objective and underperformance on another. More than half (60%) of the individual objectives included in FI XSRs that were rated low had indicators affected by missing information. This stresses the need for greater attention during project structuring and monitoring to ensuring that every approved development objective has suitable, clearly defined and measured monitoring indicators.
- 4.7 Half of the 16 corporate and infrastructure projects were rated Satisfactory or higher on effectiveness. Two such projects were rated Excellent and six Satisfactory. Five of these eight substantially achieved all of their objectives, while three failed to reach one objective but compensated for it through high achievement on the others. Four projects included targets for tax payments as an objective, but since this is a mere transfer of resources from the private sector to the public sector it is debatable whether it can be considered a true development objective. OVE did not consider increased tax payments as a development objective or rate them as such under effectiveness. Similarly, foreign exchange earnings were also mentioned as objectives in some loan documents, but these were also not considered a relevant separate development objective except in economies with currency restrictions.

17 Other examples: (i) a project's objectives included the creation of financial products with terms suitable to agricultural production cycles, without defining how the objective would be measured; and (ii) a project's objectives included financing infrastructure projects related to basic services to low-income populations, without including any indicator to measure the impact on low-income people.

4.8 All eight corporate or infrastructure projects that were rated low on effectiveness were affected by failure to achieve set objectives, and four were also affected by lack of information on some indicators. Several of the infrastructure projects suffered from shortcomings in vertical logic that limited their ability to reach the stated objectives. For example, one project claimed environmental benefits compared to road transport, when road transport was never a realistic option – the project would not have been viable with road transport; another project claimed benefits of regional integration, when it had no component contributing to that goal (which was only based on the fact that the sponsor operated in different countries).

Efficiency

4.9 Performance on efficiency across the NSG portfolio was low. Of the 43 NSG operations, 18 were rated Satisfactory and only one was rated "Excellent" on efficiency. FI operations fared worse on efficiency than the rest of the portfolio: 18 of the 27 FI operations were rated less than "Satisfactory". Two aspects are assessed under the efficiency criterion for NSG operations: project financial performance and economic performance. Of the 18 FI operations with low efficiency, four lacked information concerning both financial and economic performance. Another 12 FI operations clearly failed to reach both financial and economic performance targets. When considering only financial performance, 18 FI operations were rated less than "Satisfactory". Twelve had issues with missing information on financial efficiency indicators, while 17 also failed to reach part or all of the financial targets. In terms of economic efficiency, six FI projects were rated "Satisfactory" and three "Excellent". Of the 18 FI projects rated less than "Satisfactory" on economic efficiency, 14 had clearly not achieved economic target indicators, while seven had issues with a lack of information to assess economic efficiency.

4.10 Six of the 16 corporate and infrastructure projects were rated less than "Satisfactory" on efficiency. Two of these clearly missed both financial and economic efficiency performance targets. Four projects had clear weaknesses in the quality of information provided, partly in combination with insufficiently documented assumptions and difficult-to-follow calculations. Five of the six projects rated less than "Satisfactory" on overall efficiency were rated low on both financial and economic efficiency.

Sustainability

4.11 About two-thirds of NSG operations achieved "Satisfactory" sustainability ratings, with FI operations faring worse than the rest of the portfolio. Of the 43 validated projects, 28 had "Satisfactory" sustainability ratings; but while four-fifths of

corporate and infrastructure projects received "Satisfactory" ratings, only about half of FI operations did so. In terms of the safeguards dimension of sustainability, the XSRs presented limited information. Among the 41 XSR reports reviewed, only 13 went beyond simply reporting that the borrower had complied with safeguards requirements and set up an environment and social management system, to report more specifically on some safeguards outcomes such as the borrower's having met agreed-upon effluent standards, improved waste treatment, or reduced work-related accidents. Three reported on specifics of noncompliance. However, none of the XSRs reported against achievements under all applicable standards.

Non-core criteria

- 4.12 The majority of NSG projects performed well on additionality and investment outcome, but the quality of IDB's work was modest. Among the 43 validated projects, 34 (79%) were rated "Satisfactory" or higher on additionality ratings—considerably better than last year's results (47%)—and differences between FI and non-FI operations were small. Most NSG operations (38 of 43) were also rated "Satisfactory" or higher on investment outcome, which suggests that the IDBG did well in selecting and structuring operations that contributed positively to its financial success, and that most loans get paid as scheduled. However, the quality of IDBG's work for the reviewed operations was modest, particularly at the time of structuring and appraisal. Only one-third of the EOM 2016 operations demonstrated high work quality during structuring and appraisal as well as during supervision. Work quality was weaker for FI operations (five of 27 were rated "Satisfactory" or higher) than for corporate and infrastructure projects (10 of 16 projects).

Lessons learned

- 4.13 XSRs present lessons learned from project implementation. OVE reviewed these lessons and endorsed most of them. Where indicated, OVE suggested additional lessons as part of the validation process. These lessons are generally endorsed by Management through the comments provided to draft validations, and they become part of the knowledge system resulting from XSRs. IDB Invest reportedly has established a database system of lessons, to facilitate project teams benefitting from past lessons when preparing new operations.

B. Comparison between OVE ratings and Management self-ratings

4.14 There was considerable divergence between IDB Invest’s self-ratings in XSRs and OVE’s project outcome ratings. Using the additional information and comments provided by IDB Invest on OVE’s draft validations, OVE adjusted the ratings of some projects, but the divergence from Management’s own ratings in the XSRs initially submitted to OVE remains considerable (Figure 4.3). OVE agreed with Management’s initial ratings for 17 projects, it upgraded the overall outcome rating of three projects and downgraded 23 (Figure 4.4). These numbers refer to overall outcome ratings; most overall ratings changed by one rating step (20 of the 26 rating changes), while six projects were adjusted by two rating steps (one upgrade and five downgrades). Following further discussion with OVE, IDB Invest indicated its agreement with the final OVE ratings in all but three cases. The remaining differences in views concern for example instances in which OVE and IDB Invest agreed on shortcomings in the preparation of projects but differed in whether these should only be considered under work quality or also under relevance of design. Lack of clarity and other shortcomings in some of the model for financial and economic CBAs also contributed to ratings divergence. A joint training of OVE validators and IDB Invest staff could help limit discrepancies in the future.

Figure 4.3

Comparison between Management and OVE outcome ratings

Source: OVE.

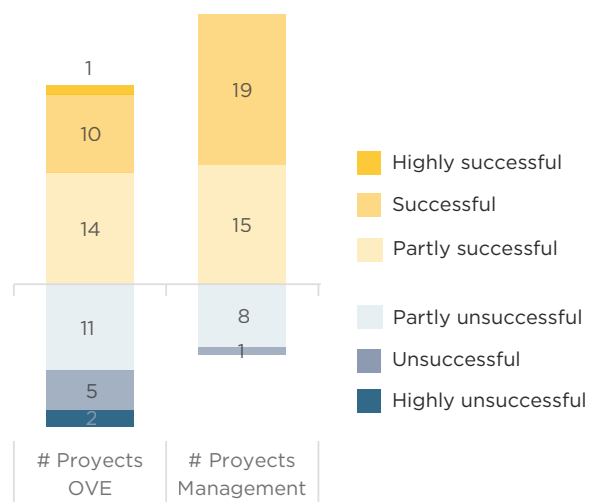
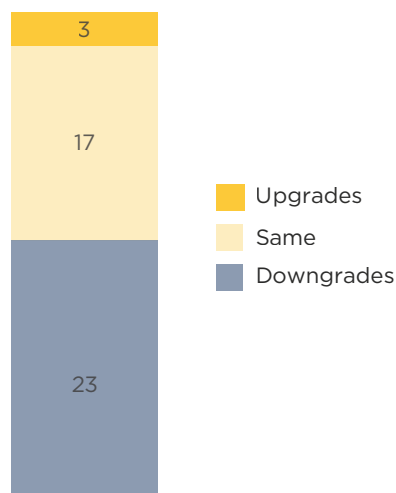


Figure 4.4
Outcome ratings changes

Source: OVE.



C. Quality of XSRs

4.15 Somewhat over half of the XSR submitted to OVE were of "Good" or "Excellent" quality. The quality of XSRs, like that of PCRs, is rated on a four-point scale, ranging from "Excellent" to "Poor". The ratings take into consideration such aspects as the XSR's clarity of presentation, internal consistency between the provided evidence and assigned rating, and the quality of financial and economic analysis. The rating also explicitly takes into account the need to seek additional clarification from the authors to carry out validations (see Annex III). The XSR quality rating is independent of results (project achievements can be rated high, while the XSR submitted for validation can receive a lower rating, and vice versa). Of the 43 XSRs reviewed, OVE classified three as "Excellent", 22 as "Good", 17 as "Fair" (at least one key shortcoming), and one as "Poor" (significant shortcomings) – see Figure 4.5. The validation process found shortcomings in the clarity of presentation, internal consistency, and in some cases, the methodological correctness of the financial and economic calculations of corporate investment projects. When OVE was unable to validate calculations because of unclear assumptions or had doubts about methodological correctness or similar concerns, OVE staff sought clarification through its Management counterparts and was then able to use improved calculations to assess a number of cases. For projects with considerable shortcomings in the calculations, OVE assigned an XSR quality rating of "Fair", provided the calculations were the only key shortcoming. Of the 16 corporate XSRs, seven had shortcomings in the calculations, and the

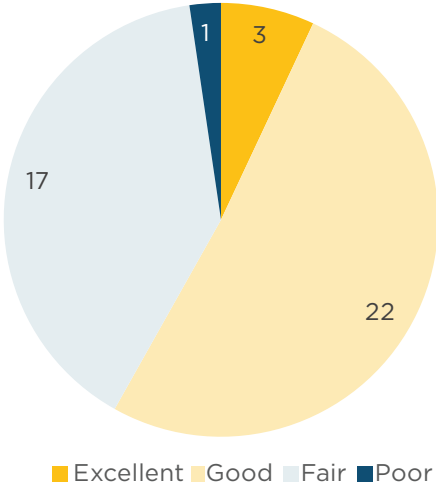
results of OVE’s revised calculations differed noticeably from those originally provided. Another three corporate XSRs also required modifications to the original calculations that did not substantially alter the calculation results. For example, for one XSR OVE used slightly modified calculations correcting inconsistencies in rate of return (ROIC and EROIC) calculations, while for another OVE adjusted the assumptions used for the calculations, informing the evaluation officer. In both cases, original targets were still met, and ratings were not affected.

4.16 In FI operations, a common issue was that OVE identified additional information in external databases or even in the IDB Invest data systems that was not included in the XSR. Validators’ needing to research additional information—for example, to determine the relevance of an operation—affected the XSR quality rating. In the 2018 validations, OVE once again upgraded ratings from “Poor” to “Fair” when a project team provided satisfactory additional information upon OVE request. However, in the interest of the efficiency of the validation process, stronger quality control of XSRs will be required before they are submitted to OVE to avoid the need for such iterations in the future.

Figure 4.5

Rating of XSR quality

Source: OVE.





05

Conclusions and recommendations

- 5.1 The IDBG has made significant progress toward a strong objectives-based self-evaluation system, and OVE's validation of self-ratings is an integral part of this system. Over the past year, significant progress has been made in consolidating the IDBG's self-evaluation system, with IDB Invest adopting updated XSR guidelines that addressed the shortcomings in the earlier pilot guidelines and IDB doing the same with PCR guidelines. IDB Invest also adopted valuable interim guidelines for the validation of FI operations. An important milestone was reached in 2018, with IDBG Management for the first time reporting OVE-validated ratings in the DEO. By adopting the validated ratings as the institutions' formal final project performance ratings, IDBG has taken a significant step toward bringing its project performance evaluation system more in line with that of other MDBs.
- 5.2 2018 is the first year during which both IDB and IDB Invest delivered self-evaluations for all operations that closed or reached EOM in 2016. IDB for the first time delivered PCRs for all operations that closed in a given year (2016), had been approved with a DEM, and required a PCR upon project completion – a significant milestone in the establishment of a credible self-evaluation system. Similarly, IDB Invest has continued its tradition of delivering XSRs for all operations that in its judgment reached EOM in 2016. An outstanding step in this respect would be to agree with OVE on a process of independent review of the project pipeline to verify the list of operations that reach EOM and are thus subject to the preparation and validation of an XSR in any given year.
- 5.3 Despite a considerable increase (by over 50%) in the numbers of PCRs and XSRs requiring validation in 2018, and intensive consultation between management teams and OVE, the joint effort succeeded in generating the validated data in time for DEO reporting. However, the number of reports being finalized and requiring validation will increase further, so management teams and OVE will need to coordinate to ensure efficient processes that allow for timely results reporting.
- 5.4 During the 2018 validation exercise, IDB Invest and OVE both applied the revised XSR guidelines, which allow for a comparison between Management's self-ratings and OVE's final ratings. This important step is expected to be reached for IDB's self-evaluations in the coming year. The 2018 PCR guidelines set a strong basis for self-evaluations, but since they were adopted only in early 2018, it was not possible for Management to prepare the PCRs for 2016 closures in accordance with them. As stipulated by the guidelines, OVE applied them to the review of the PCRs, demonstrating their application, but it is not yet possible to compare OVE's ratings directly with Management's

ratings. On IDB Invest's side, such a comparison showed considerable discrepancies between IDB Invest's initial ratings and OVE's validated ratings, with IDB Invest indicating at the end of the exercise that it finally agreed with all but 3 of 43 OVE outcome ratings.

- 5.5 The validation exercise found that over half of EOM 2016 IDB Invest operations achieved overall outcome ratings of Partly Satisfactory or higher, while somewhat less than half of the validated IDB operations that closed in 2016 did so. The lower IDB ratings are due to a combination of failure to achieve stated development objectives and lack of information on some indicators contained in the results matrix at project approval. As the 2018 PCR guidelines take hold, this issue is expected to diminish, provided project teams are made aware of the need to ensure that information contained in the original results matrix (or the matrix adopted within 60 days after eligibility) is regularly collected and finally reported on in PCRs. The validation exercise also pointed to several areas where further clarification and training of both IDBG staff and OVE validators would be beneficial to ensure consistency in evaluation and validation. Furthermore, as the number of self-evaluations is expected to increase considerably over the coming years, particularly on the IDB side, the delivery timetable and validation process will need to be further refined between Management and OVE to ensure that OVE can feasibly validate all self-evaluations in time to deliver final ratings as regular inputs to the DEO.
- 5.6 Both IDB Invest and IDB teams are establishing systems to compile lessons from closed projects. The PCR and XSR systems have increasingly become ready to generate useful lessons, based in part on improved reporting of project results and factors affecting them. Both IDB Invest and IDB management have developed or are working on systems to compile lessons to inform future operations. This is a positive outcome and a first step that facilitates the use of lessons in future projects.
- 5.7 In light of the experience of this year's validations, OVE recommends the following:

For IDB Invest Management:

- (i) In collaboration with OVE, develop guidance notes and train XSR authors and OVE validators in improved financial and economic analysis methods, to ensure common standards in terms of clearly structured calculations, homogenous methodology, and transparent assumptions.

- (ii) Agree with OVE on areas in which additional guidance notes to support the practical application of XSR guidelines would be beneficial, and develop such notes in collaboration with OVE.
- (iii) Agree with OVE on a process of independent review and verification of the project pipeline to determine the list of operations reaching EOM and subject to XSR preparation and validation every year.
- (iv) Agree with OVE on a delivery schedule of final XSRs that ensures that XSRs submitted for validation have passed all internal quality controls, taking into account the increased XSR numbers and deadlines for corporate reporting.

For IDB Management:

- (i) In collaboration with OVE, train project team leaders and PCR authors on the 2018 PCR guidelines to ensure that teams are aware of the requirements to monitor performance and report in accordance with the originally approved results matrix (adjusted within 60 days of eligibility, if applicable) and that all new PCRs are prepared according to 2018 guidelines. Make self-ratings an integral part of PCRs.
- (ii) Develop complementary sector-specific guidance notes on the preparation of CBA and CEA as an integral part of PCR preparation.
- (iii) Adopt interim guidelines for evaluating FI operations consistent with those developed for NSG projects, to bridge the period during which FI operations approved with beneficiary lists rather than sub-portfolio-specific objectives are being implemented.
- (iv) Agree with OVE on a delivery schedule of PCRs in view of increased PCR numbers and deadlines for corporate reporting.

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)