ove Office of Evaluation and Oversight

Measuring Project Performance at the IDB:

Recent Developments in the PCR and XPSR Systems

IDB
Inter-American Development Bank
June 2015

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

RE-488

TABLE OF CONTENTS

This document was prepared by a team composed of Anna Crespo (team leader), Monika Huppi, Hector Conroy, Oliver Azuara, Roni Szwedzki, Eva Bolza-Schunemann, Danya Churanek, and Kris Hallberg under the guidance of Cheryl Gray (Director). Jose Claudio Pires, Roland Michelitsch and Simon Lodato also made valuable contributions.

Acronyms and Abbreviations

CCLIP	Credit Line for Investment Projects
DEF	Development Effectiveness Framework
DEM	Development Effectiveness Matrix
DEO	Development Effectiveness Overview
ECG	Evaluation Cooperation Group
EOM	Early operating maturity
ERR	Economic rate of return
EROIC	Economic return on invested capital
FRR	Real after-tax financial rate of return
GPS	Good Practice Standards
GPS4	Fourth Edition, Good Practice Standards
IDB	Inter-American Development Bank
IDBG	IDB Group
IDB-9	IDB's Ninth General Capital Increase
IIC	Inter-American Investment Corporation
MDB	Multilateral development bank
MIF	Multilateral Investment Fund
OMJ	Opportunity for the Majority
OVE	Office of Evaluation and Oversight
PCR	Project Completion Report
PMR	Progress Monitoring Report
PSD	Private sector development
PSR	Project Supervision Report
QRR	Quality and Risk Review
ROIC	Return on invested capital
SCF	Structured and Corporate Finance Department
SPD	Office of Strategic Planning and Development Effectiveness
TC	Technical cooperation
VPC	Vice-President for Countries
XPSR	Expanded Project Supervision Report

Executive Summary

This report discusses the most recent developments in the Inter-American Development Bank (IDB, or the Bank) on measuring project performance. It has two objectives: to review the overall design of the new Project Completion Report (PCR) system for public sector operations and the Expanded Project Supervision Report (XPSR) systems for private sector operations, and to assess the most recent reports on project performance prepared by management. On the public sector side, the report reviews the 13 pilot PCRs prepared under the new guidelines and finalized by management in 2014. On the private sector side, the report presents the validation of the 2012 batch of XPSRs. They were prepared under the existing guidelines (following ECG-GPS4, the latest set of good practice standards used for private sector operations) rather than the new ones, and they include all SCF projects that reached early operating maturity in 2012.

Knowledge about what works and what does not is essential to inform decisions; therefore, the Bank has worked hard to increase its effectiveness by creating tools to enforce accountability and learning. The Bank approved the Development Effectiveness Framework (DEF) for the private sector in 2007 and for the public sector in 2008. The DEFs aimed at improving the Bank's effectiveness by "generating a body of knowledge about 'what works' in meeting the region's development challenges" (DEO 2008/09). They consist of tools to assess, monitor, and evaluate projects from their design until their completion. The IDB-9 Agreement reinforced the importance of the DEF, recommending that the Bank increase its capacity to report on results by strengthening the tools it already had in place.

When IDB projects are completed (or, on the private sector side, reach "early operating maturity"), a self-evaluation is produced: the PCR for public sector loans and the XPSR for private sector loans. These instruments are the third critical leg of the DEF; they follow up on the Development Effectiveness Matrix (DEM)/evaluability assessment and the Progress Monitoring Report (PMR)/Project Supervision Report (PSR). They are expected to report on the outcomes of projects, drawing on the same indicators and the same monitoring and evaluation plan used throughout the project cycle. As in other multilateral development banks (MDBs), the PCR and the XPSR are management's self-evaluation of project performance, typically completed by the project team and then validated by the independent evaluation office (Office of Evaluation and Oversight, or OVE).

While IDB has been at the forefront in its work on the ex-ante assessment of project evaluability, it has lagged behind other MDBs in results reporting at project closure, particularly for public sector projects. Though PCRs have existed for many years, their quality has been uneven – and generally quite low. The last PCR review, prepared by OVE in 2012, showed that project teams almost universally reported positive results from their projects, whether or not there was evidence to support these results. Interviews with staff indicated that PCRs were perceived as having limited value, getting little attention from management and receiving insufficient funding. The fact that the independent evaluation office did not review or validate PCRs also detracted from their credibility and significance. The problem of lack of evidence was not exclusive to PCRs; the same issue was identified in past OVE validations of XPSRs. In the last XPSR validation, many projects were classified as partially unsatisfactory or unsatisfactory, mostly because of the lack of appropriate evidence.

Another challenge has been that although the tools for the public and private sector have similar structures, they were designed separately and differ in many aspects. Traditionally, in most organizations public sector projects have been evaluated against their stated objectives, using the evaluation criteria of relevance, effectiveness, efficiency, and sustainability. In contrast, the development outcomes of private sector projects have been evaluated against independent benchmarks of financial performance and economic rates of return, as well as their contribution to private sector development and their environmental and social performance. Shortcomings in the current Evaluation Cooperation Group Good Practice Standards (ECG-GPS) methodology have led to discussions among ECG members, as well as within IDB, on the possible benefits of increased convergence with the evaluation standards for public sector operations.

IDB is now leading the way among MDBs in revising its PCR and XPSR systems to seek harmonization between public and private sector tools, while addressing the quality issues identified in the past. In the past two years IDB has made a major effort to completely revise the PCR and give it renewed meaning. A new system, following much more closely the ECG-GPS for public sector operations, was designed, piloted during 2013/2014, and put into operation in August 2014. The XPSR system is also being revised to assess, as in the public sector, the relevance, effectiveness, efficiency, and sustainability of interventions. A key change in the new methodology, compared to the ECG-GPS, will be to focus on the projects' objectives as the basis for the evaluation. Projects will be evaluated against the results (impacts, outcomes, outputs) they were intended to achieve. Private sector operations can focus on private sector development objectives, but can also aim to contribute more broadly to the host country's development. In addition, the XPSRs will more clearly and directly assess the relevance of the operation ex-post and the attribution of observed outcomes to IDBG activities. At the same time, efforts have been made to ensure that the particularities of private sector operations will be considered.

OVE's review of the PCR pilot phase found that the overall quality of the reports was much higher than earlier PCRs, but there are still shortcomings. The analyses produced under the PCR pilot were generally supported by data, and a clear effort was made to better document results achieved. Most problems identified by OVE arose from shortcomings in the guidelines, most notably an inflexible rating system and inadequate space to discuss issues of implementation or ex-post problems with project design. Because low evaluability also affected the quality of PCRs for older projects, management is applying the new PCR system only to projects approved after 2009.

A main drawback of the current PCR rating system is that it diverges from the standard approach described in the ECG-GPS and used by all IDB private sector windows and other MDBs, thus hindering comparability across the IDB Group and with other organizations. While the standard approach utilizes a discrete 4- or 6-point rating system divided evenly between positive and negative ratings, IDB's PCR guidelines use a continuous numerical rating system for effectiveness and overall performance and a 5-point scale for other core criteria.

For the XPSRs reviewed, the ratings were generally positive, but the quality of the XPSRs could be further improved. Most XPSRs did not document well the effectiveness of projects, in part because of limitations in GPS4. They often did not discuss the results chain, hindering the connection between claimed outcomes and project activities. Strengthening the discussion of the causal chain and the evidence on effectiveness, in line with the new guidelines, should improve the quality of the XPSRs and strengthen the lessons learned.

Looking forward, OVE has the following recommendations:

1. Revise the PCR guidelines to further harmonize them with those for the private sector and to address some shortcomings identified in this report. In particular:

    a) Revise the rating system. OVE recommends that management work in partnership with the private sector windows to harmonize their rating systems, using a 4-point scale for all core criteria and a 6-point scale to assess the overall performance of the projects. The current system of continuous numerical ratings could be used as input but would not constitute the final ratings reported for projects. In validating management's project ratings, OVE would take into consideration any numerical ratings but would also use discretion to arrive at what it considers to be the most sensible final project ratings.

    b) Include space for discussion of implementation progress and an ex-post analysis of project design. One of the main weaknesses of the new guidelines is the omission of discussion on implementation progress and project design, which also limits the discussion of lessons learned.

    c) Include compliance with environmental and social safeguards as a factor in sustainability. As in the private sector guidelines, PCRs should consider compliance with safeguards as a dimension of sustainability rather than have it as a separate non-core criteria as in the current PCR guidelines.

2. Continue the process of implementing the new XPSR guidelines to improve SCF's and OMJ's ability to report on results, and to further harmonize with the other IDB windows. As the Bank moves towards the merger of all private sector windows into IIC, it is more important than ever that the private sector windows harmonize their systems to

report on results. The new methodology has a stronger focus on development results and is more harmonized with the public sector methodology than the previous system, and it will enhance the rigor with which information is presented. It will further reinforce the need for projects to be evaluable at approval.

I. Introduction

1.1 **Knowledge about what works and what does not is essential to inform decisions by and, consequently, increase the effectiveness of any development organization.** For this reason, the international development community has increasingly focused its attention on developing and putting in place instruments to measure the effectiveness of its work. Better knowledge about the interventions and their results leads to more concrete and precise recommendations about how to work better in the future. The aim is to ensure that the resources allocated to development are used as effectively as possible, leading to tangible and sustained improvement in the lives of the beneficiaries.

1.2 **This report reviews recent developments in the project-level reporting tools used by the Inter-American Development Bank (IDB, or the Bank), and recent project results assessed through these tools.** On the public sector side, this report assesses the new methodology for preparing Project Completion Reports (PCRs) and presents the main findings of the pilot exercise concluded in 2014. On the private sector side, the report describes progress in developing new guidelines for reporting on results, and it discusses the results of the sixth exercise to validate Expanded Project Supervision Reports (XPSRs) for Bank private sector projects that achieved early operating maturity[1] (EOM) in 2012, before the change in methodology.[2]

1.3 **As background, both the PCRs and the XPSRs are part of IDB's broader Development Effectiveness Framework (DEF), finalized for the private sector in 2007 and for the public sector in 2008** (Box 1.1). The DEF is intended to improve the Bank's effectiveness by "generating a body of knowledge about 'what works' in meeting the region's development challenges" (Development Effectiveness Overview, DEO 2008/09). It consists of tools to assess, monitor, and evaluate projects from their design until their completion.[3] One of the main challenges in

---

[1] EOM occurs when operations meet a set of objective criteria defined by the Good Practice Standards (GPS). Corporate projects reach EOM when (a) the project financed has been substantially completed, (b) the project financed has generated at least 18 months of operating revenues for the company, and (c) the Bank has received at least one set of audited annual financial statements covering at least 12 months of operating revenues generated by the project. By contrast, the GPS defines EOM for financial projects as being at least 30 months after the Bank's final material disbursement for subloans or subinvestments, i.e., ignoring disbursements for small follow-up investments in existing client companies and disbursements to cover management fees or other expenses of the investment funds themselves.

[2] This exercise was carried out in line with the fourth edition of the Good Practice Standards (GPS4) for Private Sector Operations issued by the Multilateral Development Banks' (MDBs') Evaluation Cooperation Group (ECG). The GPS are available at www.ecgnet.org. The new guidelines are applied to projects that achieved EOM in 2013 and 2014.

[3] The importance of this subject is reflected in many of the requirements in the IDB's 9th General Capital Increase (IDB-9). The IDB-9 Agreement focuses on the need to increase

using the DEFs up to now has been that although the tools for public and private sector have similar structures, they were designed separately and differ in many aspects.[4] As will be discsussed below, the IDB is working to unify these approaches.

Box 1.1 Quick Summary of the Public and Private Sector DEFs

The public sector DEF requires each project to be designed, monitored, and evaluated with three instruments:

- Development Effectiveness Matrix (DEM): prepared during design to assess the project's ability to report on results at completion;
- Progress Monitoring Report (PMR): prepared twice a year throughout the implementation of the project to monitor outputs and costs defined in the Results Matrix; and
- Project Completion Report (PCR): prepared at project completion to show evidence of the achievements of the project vis-à-vis its intended objectives, and to discuss the sustainability of those achievements.

The private sector framework also comprises three tools, which are analogous to the ones described for the public sector but somewhat different in content:

- Evaluability Assessment (evaluability score) and DE project assessment (DEM score);
- Project Supervision Report (PSR), which updates the DEM score and the results matrix and is prepared annualy; and
- Expanded Project Supervision Report (XPSR), produced for all projects that reach EOM.

1.4 **When IDB public sector projects are completed or private sector projects reach EOM, the DEF requires that IDB management produce a self-evaluation: The Project Completion Report (PCR) for public sector loans and the Expanded Project Supervision Report (XPSR) for private sector loans.** These instruments report on the outcomes of projects, drawing on the same indicators and monitoring and evaluation plan used throughout the project cycle. According to the ECG-GPS, such self-evaluations should be validated by the organization's independent evaluation office.

---

the Bank's capacity to report on results: "We recognize that the Bank must not simply become larger, but that it must also become more effective at achieving its mandates. Thus, the Board of Governors agrees to the following reforms: … improving the results measurement framework to identify for every project whether specific and tangible results have been achieved" (CS-3868-1).

4 The ECG was set up to share knowledge and, to the extent practical, harmonize approaches among the independent evaluation offices of nine MDBs and the IMF. Since 2000, the ECG has set common metrics for project evaluation that ECG members have striven to apply, formalized in the GPS. The GPS, used by most MDBs, evaluate public sector projects against their stated objectives using the evaluation criteria of relevance, effectiveness, efficiency, and sustainability. In contrast, the development outcomes of private sector projects are evaluated against independent benchmarks of financial performance and economic rates of return, as well as their contributions to private sector development and their environmental and social performance. Shortcomings in the current ECG-GPS methodology have led to discussions among ECG members, as well as within IDB, on the possible benefits of increased convergence between the evaluation standards for public and private sector operations.

1.5 **While IDB has been at the forefront in its work on the ex-ante assessment of project evaluability (the DEM), it has lagged behind other multilateral development banks (MDBs) in results reporting at project closure, particularly for the public sector.** Though PCRs have existed for many years, their quality has been uneven – and generally quite low. The last PCR review, prepared by the Office of Evaluation and Oversight (OVE) in 2012, showed that project teams almost universally reported positive results from their projects, whether or not there was evidence to support these results.[5] Interviews with staff indicated that PCRs were perceived as having limited value, getting little attention from management and receiving insufficient funding.[6] The problem of lack of evidence was not exclusive to PCRs; the same issue was identified in past OVE validations of XPSRs. In the last XPSR validation many projects were classified as partially unsatisfactory or unsatisfactory, mostly because of the lack of appropriate evidence.

1.6 **IDB is now leading the way among MDBs in revising its PCR and XPSR systems to seek harmonization between public and private sector tools, while also addressing the quality issues.** In the past two years IDB has made a major effort to completely revise the PCR instrument and give it renewed meaning. A new system was designed, piloted, and put into operation in August 2014.[7] It is expected that earlier efforts on the implementation of the DEM and the PMR will lead to better-quality PCRs over time. The XPSR system is also being revised. Like the public sector PCRs, the new XPSR guidelines focus on the achievement of project objectives—asking whether those objectives are relevant to the country's development needs and whether the project achieves those objectives effectively, efficiently, and sustainably. At the same time, efforts have been made to ensure that the particularities of private sector operations will be considered, and characteristics of the GPS for Private Sector Operations are being kept in the guidelines. OVE is now working with the Development Effectiveness Units of all IDB Group private sector windows—SCF, OMJ, MIF, and IIC—to develop new guidelines for aligning their XPSRs among themselves and with the IDB public sector PCRs.

5 The most recent review was OVE's *Review of the Project Completion Reporting System for Sovereign Guarantee Operations* (RE-417), July 2012, available at www.iadb.org/evaluation.

6 While in the past OVE validated XPSRs, it did not regularly review or validate PCRs in the past, which also detracted from their credibility and significance.

7 The new approach closely follows the ECG-GPS, focusing on the achievement of project objectives and assigning ratings for relevance, effectiveness, efficiency, and sustainability, as well as an overall project rating. IDB does not plan to rate Bank or borrower performance, which are two additional "below the line" criteria included in the GPS.

II. Public Sector Project Evaluation

A. The New PCR System

2.1 **In 2013, a working group composed of representatives of the Vice-President for Countries (VPC), Vice-President for Sectors (VPS) and OVE, and headed by the Office of Strategic Planning and Development Effectiveness (SPD), was formed to develop new guidelines for PCRs.** The goal was not only to seek harmonization with the GPS and other tools in the DEF, but also to strengthen the incentives for the production of good-quality PCRs. The new guidelines were ready for piloting in early 2014.

Figure 2.1 Old PCR versus the New PCR Core Criteria

Old PCR Criteria

Development Objectives assesses the difference between achieved and planned outcome indicators. In case of a significant gap between them, a brief explanation about the factors responsible for the gap was required. In addition, the inclusion of estimations of the internal rate of return and of cost-effectiveness for all PCRs of projects that presented such analysis at approval was required.

Implementation Progress is an analysis of delivery of outputs, considering the delivery schedules and quality of the outputs delivered.

Sustainability takes into consideration (i) analysis of the critical factors related to the sustainability of the results achieved throughout project execution, including a future operation plan; (ii) potential risks that could affect sustainability of results and how these risks will be managed, including recommendations for Bank follow-up actions; and (iii) EA/borrower's capacity to sustain project results.

New PCR Criteria

Relevance refers to the consistency of the project's objectives with beneficiary needs, the country's development or policy priorities and strategy, and the Bank's assistance strategy and corporate goals. Relevance is assessed against circumstances prevailing at the time of the evaluation.

Effectiveness evaluates the extent to which the project achieved, or is expected to achieve, its stated objectives.

Efficiency is assessed by one of two dimensions: (i) extent to which the benefits of the project (achieved or expected to be achieved) exceed project costs; or (ii) extent to which the benefits of the project were achieved at less than expected or at reasonable cost.

Sustainability assesses the probability and the impact of various threats to the continuation of outcomes beyond the completion of the project (ex-post project implementation), taking into account how these threats have been mitigated in the project's design or during execution.

2.2 **The new PCR guidelines fundamentally change the criteria upon which a project is assessed.** The old PCR methodology called for three evaluation criteria to be assessed and rated: Achievement of Development Objectives, Implementation Progress, and Sustainability. The new criteria, which are more consistent with the GPS, call for four "core" evaluation criteria to be assessed and rated: Relevance, Effectiveness (achievement of objectives), Efficiency, and Sustainability.[8] Figure 2.1 compares the old and new PCR core criteria.

2.3 **There are also non-core criteria in the new guidelines that follow the DEM and are not rated in the PCR:** the project's contribution to the Bank's corporate strategic development objectives, the project's contribution to the development objectives of the country strategy, the quality of implementation of the project's monitoring and evaluation plan, the use of country systems, and the implementation of social and environmental safeguards in the project.

2.4 **Although the new PCR guidelines are closely aligned with the GPS in terms of core criteria, the rating system used in IDB differs considerably from what is used in other MDBs and from the ECG-GPS.** IDB's guidelines rate project effectiveness and overall project outcome using a continuous numerical rating scale from 1 to 100, and they rate efficiency, relevance, and sustainability using a discrete 5-point scale.[9] In contrast, other MDBs follow the ECG-GPS in using either a 4- or 6-point scale for all core criteria, with qualitative labels evenly distributed between positive and negative. The GPS advocates an even-numbered scale because it forces evaluators to recognize a positive or negative tendency, while an odd-numbered scale allows evaluators to select a more neutral middle option. The GPS includes qualitative labels because they help ensure a common interpretation of each rating, possibly reducing measurement error. In addition, IDB's assessment of overall performance brings all core criteria together using a pre-defined formula, while other MDBs allow some discretion in how various core criteria are weighed in each case.

2.5 **The Bank has also revised the process for preparing PCRs.** The new guidelines define the timeframe for report preparation very clearly. Generally, project teams should have more time to prepare each PCR under the new guidelines. The Quality and Risk Review (QRR) will remain virtual,[10] but chaired by the VPC manager in an attempt to give

---

8 Of the three core criteria of the old system, only Sustainability was the same as the GPS core criterion. Development Objectives corresponded roughly to the Effectiveness criterion in the GPS, but the discussion of outcomes was separated from the discussion of outputs, splitting the discussion of the results chain into two parts. Development Objectives also included some elements of project efficiency. The third PCR criterion, Implementation Progress, did not correspond to a GPS criterion, although it could usefully be included in the effectiveness section of the GPS as part of the discussion of the causal chain.

9 In the new PCR checklist, where the ratings system is defined, the 5-point scale is labeled as Full Achievement, High Achievement, Partial Achievement, Low Achievement, and No Achievement, from the highest to the lowest number. However, these scales do not assess effectiveness, but efficiency, relevance, and sustainability, so the use of the term achievement is unclear.

10 According to the new guidelines for PCR preparation, "Chairperson may decide to have a face to face QRR after request of any of the QRR members. For a face to face QRR, the revision period lasts no longer than five (5) working days after distribution of the PCR. The face to face meeting must take place no later than ten (10) working days after distribution. In case of a face to face QRR, the PCR team leader presents the PCR.

the peer review process more importance and attention than in the past.[11] After the comments of the QRR have been taken into consideration, the VPC manager must approve the revised PCR. The consultation with government follows, as in the prior system, and the final version of the PCR must be approved by the division chief and VPC manager. OVE plans to validate a sample of PCRs annually, reporting a summary of project results in its annual report and in the DEO. All PCRs will continue to be disclosed to the public. The goal is to give the PCRs a higher level of visibility and promote accurate assessments and ratings of projects results.

B. The PCR Pilot

2.6 **Thirteen projects were selected to participate in the pilot exercise, though one PCR, GY-0011, was never concluded.** The sample included projects from different divisions for which a PCR was due. The projects fall into three of the Bank's five institutional priorities:[12] (i) Social Policy for Equity and Productivity: CH-L1014, CO-L1059, CO-L1010, HA-L1062, ME-L1080; (ii) Infrastructure for Competitiveness and Social Welfare: AR0202, PE-L1121, NI-L1006, HO-L1007; and (iii) Institutions for Growth and Social Welfare: AR-L1008, BR-L1104, and PR-L1032. They were approved between 2000 and 2012, and five of them had a DEM.[13] The sample includes one policy-based loan (PE-L1121), and three projects in the sample belong to multiphase operations (CH-L1014, CO-L1059, and PE-L1121). Table 2.1 provides detailed information on the sample.

2.7 **Because the rating system, as reported in the 2014 DEO, has changed as a result of the pilot, OVE decided to rate the projects according to the newest rating system (in the guidelines in effect as of August 2014) rather than validate the pilot ratings.** OVE discussed its draft ratings with PCR team leaders before finalizing them. It is important to point out that the ratings under the original system and in

---

During the QRR, SPD/SDV provides a preliminary validation of the ratings prepared by the team to each of the PCR components, based on compliance with completion reporting policy and standards. The preliminary validation is conducted using the information of the PCR that was sent to QRR."

11 OVE's 2012 PCR review pointed out that the QRR, which in the old system was expected to be chaired by the division chief, was rarely helpful in improving the quality of the PCR. "The analysis of available QRR information revealed that in about 25% of the cases the QRR was attached but it generated no comments. Moreover, only slightly more than half of the survey respondents said that the comments received at the QRR meeting were useful. Only about a third of respondents said that QRR comments addressed the findings, conclusions, and recommendations in the PCR or important topics that had been overlooked. Problems with data or the empirical basis of conclusions drawn in the PCR were covered only 23 percent of the time."

12 The other two are *Competitive Regional and Global International Integration* and *Protecting the Environment, responding to Climate Change, Promoting Renewable Energy and Enhancing Food Security*. For more information see AB-2764, Annex I "IDB Results Framework 2012-2015."

13 The DEM was introduced in 2009.

OVE's assessment following the system in effect in August 2014 are highly correlated.[14]

2.8 **The PCRs were rated according to the four core criteria – effectiveness, efficiency, relevance and sustainability – which together automatically resulted in the overall performance rating.** OVE followed the PCR checklist and took note of any difficulty in applying ratings to the projects. The rating on effectiveness could assume a value between 0 and 1, depending on the percentage of the target achieved for each indicator. For outcome indicators, only those with evidence of attribution to the project were considered. Higher weight was given in the overall effectiveness assessment to attributable outcomes achieved (60%) than to outputs achieved (40%). Efficiency (Table 2.2), relevance (Table 2.3), and sustainability (Table 2.4) were rated on a five-point scale, with the values 0, 0.25, 0.5, 0.75 and 1. In line with the ratings for evaluability in the DEM, the overall performance rating was defined as a weighted average of the ratings on the four core criteria, with effectiveness counting 40%, efficiency 30%, relevance 20%, and sustainability 10%.

14 The ratings reported in the DEO 2014 are those prepared during the pilot, and therefore never validated by OVE. OVE's ratings presented in this report are numerically very close to the ones in the DEO, although they follow a revised rating system.

Table 2.1 Summary of Projects in Pilot PCR Exercise

Project number	Project name	Type	Approval date	Project objective	Approved amount in US$m	Total costs in US$m
AR-L1008	Program for the Institutional Strengthening of the Argentine Senate	Investment Loan	2004	Institutional strengthening of the Argentine Senate by improving its parliamentary management and lawmaking processes; administrative, financial, and human resource management; optimization and training; as well as change management and distribution of public information.	4.8	8
AR-L0202	Improvement of Border Crossings and Integration of Traffic Corridors.	Investment Loan	2000	Facilitate passenger and cargo transportation in priority corridors of the national road network (especially with Chile and the broader Mercosur). Reduce transport costs and travel times, improve and provide alternative routes for traffic, exports and imports through the ports of the Atlantic and the Pacific Coast, and increase security for the users of those routes.	200	400
BR-L1104	PROCIDADES for the Municipality of Campo Grande Integrated Development Program	Investment Loan	2008	Enhance the quality of life of residents of the Municipal of Campo Grande by implementing urban projects and actions to improve city government workings. Revitalize downtown Campo Grande; increase mobility across the city and the administrative efficiency of the city government.	19.4	46.6
CH-L1014	Programa Orígenes, Phase II: Integral Development Program for Indigenous Peoples	Multi-Phase Program Loan	2006	Improve the living conditions of 1000 rural indigenous communities in five regions, based on respect and consideration for their cultures, with ample opportunity for horizontal participation and an exchange of know-how between indigenous peoples and public services.	45.2 (Phase II) 80 (Total IDB)	109 (Phase II) 133.4 (Total)
CO-L1010	Bogota Equity in Education Program	Investment Loan	2006	Support the implementation of the District of Bogota's strategy to enhance equity and quality in preschool, primary, secondary, and postsecondary education.	60	90
CO-L1059	*Familias en Accion* Conditional Cash Transfer Program - Phase II	Multi-Phase Program Loan	2010	Supplement the income of poor urban families with children under 18 years; promote human capital formation by increasing the use of early childhood development, educational, and health care services; and improve the program's quality through strengthening its technical management tools and evaluation agenda under way.	220 (Phase II) 526 (Total IDB)	220 (Phase II) 1267 (Total)

HA-L1062	Emergency Response for the Containment of Cholera	Investment Loan/ Investment Grant	2010	Contribute to the Government's Inter-Sector Response Strategy to reduce morbidity and mortality due to the cholera outbreak in Port-au-Prince's poorest neighborhoods by strengthening the ministry's stewardship role, as well as the national water and sanitation company's response capacity and surveillance of the quality of water.	15 (+5m IDB grant)	20
HO-L1007	Low-income Housing Program	Investment Loan	2006	Improve the housing and habitat conditions of low- and middle-income families. Enable the government to meet multiple demands in the housing sector, giving families greater access to formal housing and basic urban services.	30	31
ME-L1080	Strengthening the Human Development Program *Oportunidades*	CCLIP	2010	Reinforce the process of adjusting the urban operating model for Oportunidades, based on the operational evaluation of the initial phase of that process, to contribute to ending the intergenerational transmission of poverty.	800	800
NI-L1006	Acopaya-San Carlos-Costa Rican Border Highway Integration Program under the Puebla-Panama Plan.	Investment Loan	2006	Make Nicaragua more competitive by further integrating the Chontales and San Juan River regions into the national economy and Meso-American region. Achieve a highway connection between Acoyapa, San Carlos, and the Costa Rican border that will allow constant and safe traffic with lower transportation costs and shorter travel times.	49.5	60.7
PE-L1121	Development of a New Sustainable Energy Matrix for Peru - NUMES	Programmatic Policy-Based Loan	2012	Support the development of a New Sustainable Energy Matrix to maximize the benefit derived from energy resources in a sustainable manner. This is the fourth operation in a series of policy-based loans, whose objective is to provide continuity of support for the policy reforms, sector decisions, the formulation of subsector plans and implementation of other specific measures.	30 (Phase IV) 230 (Total IDB)	30 (Phase IV) 230 (Total)
PR-L1032	CCLIP to provide financing to the *Agencia Financiera de Desarrollo*.	CCLIP	2009	Continue helping Paraguay's productive sector to gradually become more competitive by providing medium- and long-term financing for business ventures.	50 (Phase II) 150 (Total IDB)	50 (Phase II) 150 (Total)

Table 2.2 Efficiency Rating Criteria

Value	Definition
1	If CBA was conducted: ex-post ERR exceeds by 10% or more the ex-ante ERR. If CEA was conducted: the project results were achieved at no more than 90% of the cost of the alternative ways to achieve the same or similar results.
0.75	If CBA was conducted: ex-post ERR exceeds by 5% or more the ex-ante ERR. If CEA was conducted: the project results were achieved at no more than 95% of the cost of the alternative ways to achieve the same or similar results.
0.5	If CBA was conducted: ex-post ERR exceeds (or equals) the ex-ante ERR. If CEA was conducted: the project results were achieved at a cost below (or equal to) the cost of the alternative ways to achieve the same or similar results. If CTOA was conducted: the PCR reports that the project is classified as "Satisfactory" in the PMR (in relation with the CPI or SPI indexes, whichever shows the largest problem).
0.25	If CBA was conducted: ex-post ERR is below the ex-ante ERR. If CEA was conducted: the project results were achieved at a cost above the cost of the alternative ways to achieve the same or similar results. If CTOA was conducted: the PCR reports that the project is classified as "Alert" in the PMR (in relation with the CPI or SPI indexes, whichever shows the largest problem).
0	If CBA was conducted: ex-post ERR is below the ex-ante ERR by more than 5%. If CEA was conducted: the project results were achieved at a cost 5% (or more) above the cost of the alternative ways to achieve the same or similar results. If CTOA was conducted: the PCR reports that the project is classified as "Problem" in the PMR (in relation with the CPI or SPI indexes, whichever shows the largest problem).

Notes: CEA stands for Cost Effectiveness Analysis; CBA is Cost Benefit Analysis; CTOA is Cost and Time Overrun Analysis; ERR is Economic Rate of Return; CPI is Cost Performance Index, and SPI is Schedule Performance Index. For projects lacking CEA or CBA, a maximum score of 0.5 (Partial Achievement) can be assigned. This criterion is assessed but not rated for policy-based loans.

Table 2.3 Relevance Rating Criteria

Value	Definition
1	The conditions that made the project relevant at the time of approval remain the same, or the conditions that made the project relevant at the time of approval are not the same but the Bank fully accommodated to the changing needs of the client during the period of implementation.
0.75	The conditions that made the project relevant at the time of approval are not the same, but the Bank highly accommodated to the changing needs of the client during the period of implementation.
0.5	The conditions that made the project relevant at the time of approval are not the same, but the Bank partially accommodated to the changing needs of the client during the period of implementation.
0.25	The conditions that made the project relevant at the time of approval are not the same, but the Bank poorly accommodated to the changing needs of the client during the period of implementation.
0	The conditions that made the project relevant at the time of approval are not the same, and the Bank was not responsive to the changing needs of the client during the period of implementation. Or, the conditions that made the project relevant at the time of approval are the same but the assessment of relevance at entry was poorly conducted, the real project relevance was always limited, and the Bank was not responsive to adapt the project to the needs of the client.

Note: Relevance (alignment and contribution of the project to the needs of the beneficiaries, the priorities of the country and the Bank) is assessed against circumstances prevailing at the time of the evaluation. When possible, the Relevance assessment refers back to the ex-ante assessment of Strategic

Alignment in the DEM. If circumstances significantly changed during project implementation, then the Relevance assessment seeks to assess whether the Bank's execution assistance was responsive to changing needs during the full period of implementation and whether the operation remained important to achieving country, Bank, regional, and global development objectives.

Table 2.4 Sustainability Rating Criteria

Value	Definition
1	There are no significant risks to the sustainability of current and future outcomes, or 100% of the existing risks can be mitigated.
0.75	There are significant risks to the sustainability of current and future outcomes, but more than 75% of the existing risks can be mitigated.
0.5	There are significant risks to the sustainability of current and future outcomes, but more than half of the existing risks can be mitigated.
0.25	There are significant risks to the sustainability of current and future outcomes, but more than 25% of the existing risks can be mitigated.
0	There are significant risks to the sustainability of current and future outcomes, and less than 25% of the existing risks can be mitigated.

2.9 **The overall quality of the PCRs in the pilot is significantly higher than the quality of PCRs produced under the old guidelines.** OVE's 2012 report found that the ratings for achievement of objectives had no correlation with the evidence presented.[15] In contrast, OVE assessed the quality of the information provided in the PCRs in the pilot, and it rated only one PCR as unsatisfactory. This was PR-L1032, which aimed at helping Paraguay's productive sector to gradually become more competitive by providing medium- and long-term financing for business ventures. However, it is important to note that this project is part of a Conditional Credit Line for Investment Projects (CCLIP), which is not yet completed, and therefore it is unclear why it is part of the sample.[16] In addition to better evidence, all of the pilot PCRs included a discussion of indicators and of any changes made to the results matrix.

2.10 **OVE's scores on the various criteria are shown in Table 2.5.** One of the main problems identified in the past was the fact that almost all PCRs were rated identically, as satisfactory. Under the new methodology the overall performance of the projects varies from 0.36 to 0.87, on a scale from 0 to 1.

---

15 OVE RE-247-2, RE-315, and RE-417, which states: "The Bank's self-assessments endorse PCR ratings without requiring documented evidence," "too much outcome information is missing to assess development effectiveness," and "The lack of outcome evidence constrains Bank's capacity to manage for results."

16 According to the 2014 PCR guidelines: "The basic unit of accountability is either an operation or a series of operations with a common set of objectives that were approved by the Board. This means that operations in a series that share a common set of objectives, such as supplemental loans, programmatic PBLs, MPPs, and CCLIPs, are evaluated in a single PCR prepared for the final operation in the series. This series is assessed as a whole."

Table 2.5 PCR Core Criteria Score Assigned by OVE Summary

	Effectiveness	Efficiency	Relevance	Sustainability	Overall performance
AR0202	0.65	0.25	1.00	0.75	0.61
AR-L1008	0.66	0.25	1.00	0.75	0.61
BR-L1004	0.93	0.75	1.00	0.75	0.87
CH-L1014	n.a.[(a)]	0.75	1.00	0.50	n.a.[(a)]
CO-L1010	0.73	0.75	1.00	0.75	0.79
CO-L1059	0.51	0.75	1.00	0.75	0.70
HA-L1062	0.91	0.25	1.00	0.75	0.71
HO-L1007	0.95	0.25	1.00	0.75	0.73
ME-L1080	1.00	0.50	1.00	0.50	0.80
NI-L1006	0.67	0.00	1.00	0.75	0.54
PE-L1121	0.93	n.a.[(b)]	1.00	0.75	n.a.[(b)]
PR-L1032	0.40	0.00	0.75	0.50	0.36

Note: [(a)] CH-L1014 is the oldest project in the sample. The first operation was approved in 2001. The results matrix of this project has changed over time, and no target was set to the indicators tracked in the PCR; therefore the formulas for effectiveness and the overall performance could not be applied.

[(b)] PE-L1121 does not have a cost-benefit analysis, as the effects of the program are expected to be observable only in the long run. The PCR claims that an analysis will be performed in 2017. Without the ex-post economic analysis it was not possible to assign a rating, nor to compute the overall performance rating for this project.

2.11 **OVE was unable to apply the new rating system to two projects, CH-L1014 and PE-L1121.** Since the ratings are mostly defined through arithmetic formulas, whenever information is missing, ratings cannot be assigned. For CH-L1014, the project results matrix was reconstructed during project implementation and no targets had been assigned, so the formula could not be applied. For PE-L1121, it was not possible to compare the ex-ante cost benefit analysis with the ex-post, since the economic analysis of this project will only be conducted in a couple of years. As a consequence, no rating could be assigned.

2.12 **Most of the variance among projects is in effectiveness, for which scores depend on the achievement of outputs and outcomes.** According to OVE's assessment, five projects had satisfactory or highly satisfactory achievement of objectives. The two projects with lowest achievement were CO-L1059, an education project in Colombia, and PR-L1032, noted above. CO-L1059 supported the implementation of Bogota's strategy to enhance equity and quality in pre-school, primary, secondary, and postsecondary education. Only 16 of the 20 schools initially planned were built, though the quality of these schools was satisfactory. In addition, many indicators in the PCR could not be attributed to the project alone, since the program it was supporting became a national program, going beyond Bogota.

2.13 **The ratings on effectiveness are highly dependent on the quality of the indicators used.** All PCRs included data on the indicators in the results matrix; however, not all these indicators were appropriate measures to assess the effectiveness of the project. For instance, PR-L1032, which supported AFD (Agencia Financiera de Desarrollo), defined output rather than outcome indicators, and this resulted in a low effectiveness rating. However, OVE's intermediate assessment of this operation in the Paraguay Country Program Evaluation, which went beyond the indicators proposed, says that "This institution [AFD] has also become a key instrument for the government to direct credit to the private sector and administer new funds from Itaipú. In and of itself, the consolidation of the AFD is an accomplishment attributable to the Bank's timely and flexible intervention. In terms of the results of IDB funding for the AFD, the resources were placed at an average of four times the system's average (nine years), with low default rates (0.372%), in the housing (38%) and productive sectors (62%)." This points to a potentially higher level of effectiveness than suggested by the rating attributed in the PCR.

2.14 **The ratings on efficiency were generally low in the pilot, mostly because of the lack of ex-ante cost-benefit or cost-effectiveness analysis.** Half of the projects in the pilot did not have a cost-benefit or cost-effectiveness analysis. Two had a cost and time over-run analysis, and two others provided a discussion on the costs of the project. This section of the PCRs was generally the weakest one, with only four projects presenting a solid analysis. This may affect many more PCRs to come, as an ex-ante economic analysis has been required only since 2011.

2.15 **Relevance ex-post remains high and is highly correlated with relevance ex-ante, since there was no evidence that the projects deviated from the Banks' or the countries' strategic objectives.** Only one project, PR-L1032, scored below 1. The rating followed the criteria agreed with SPD, based on the evidence provided and considering that the CCLIP is not yet concluded.

2.16 **The rating on sustainability considers the identification of risks and mitigation measures.** Given that evaluators doing desk-based validations are not generally able to identify risks that are not included in project documents, the analysis here takes as face value the risks identified and considers the analysis presented. In many cases identified risks, particularly political risks, did not have associated mitigation measures, or likely impacts were not fully discussed in the PCRs.

C. OVE's Assessment of the New PCR Methodology

2.17 **OVE finds two main problems in the new PCR methodology. The first relates to the rating system, which in OVE's view has four major disadvantages.**

(i) *Obscurity.* The use of a continuous numeric rating system on a scale from 0 to 1 obscures the meaning of the ratings – for example, it is not obvious what the difference is in practice between a rating of 0.87 and 0.61.

(ii) *Inflexibility.* In an attempt to increase the rigor with which ratings are assigned, the new system has become overly inflexible and mechanical, and at times inapplicable. For example, all project objectives and components must be weighted equally, often leading to misleading results (e.g., not allowing for the differential importance of different components). In BR-L1104 (PROCIDADES Campo Grande), for example, the outcomes of the numerous infrastructure components must all be given the same weight, and also the same weight as the outcome of the smaller institutional capacity component. Another example of inflexibility is in the way the efficiency rating is calculated, as it does not compare project economic rates of return (ERRs) to an absolute benchmark, but rather compares only ex-ante and ex-post project calculations, potentially leading to misleading assessments.[17] Another issue with the efficiency rating is what to do with emergency loans. Because such loans are not normal operations and are prepared quickly to address a country's urgent need, the rigor of the ex-ante economic analysis is usually lower.[18]

(iii) *High dependency on the quality of project design.* Projects with evaluability issues at design might not be rateable at all. Since ratings are defined though formulas, a rating cannot be assigned for the project as a whole if data have not been collected or targets not defined ex-ante for some indicators. For example, as mentioned above, the indicators for CH-L1014 were appropriately changed during implementation, but no targets were defined for them. In this case the formula for assessing effectiveness is not applicable, and OVE left the rating as "n.a."[19]

(iv) *Lack of comparability across the IDB Group.* The continuous numerical rating approach in the PCR guidelines is not comparable with the 4- and 6-point scales used in the proposed XPSR guidelines, making it impossible to compare results achieved across the IDB Group.

---

17 For example, in NI-L1006, the ex-ante and ex-post ERR for this project were properly calculated, and by the standards of the sector both ERRs are high and do not characterize an inefficient project. In this regard, the PCR states: "A partir de los resultados de los análisis costo beneficio ex-ante y ex-post se concluye que las inversiones realizadas en las obras viales han sido eficientes y beneficiosas para la sociedad." However, despite the absolute numbers, since the ex-post ERR (13.4%) is more than 5% lower than the ex-ante (20%), the project got a low rating in efficiency.

18 HA-L1062 was an emergency project to reduce morbidity and mortality due to the cholera outbreak in Port-au-Prince. The efficiency rating of this project under the current guidelines was very low, given the limitation in data. However, given the nature of the project, not much could be expected in terms of analysis of efficiency, which should arguably be considered in the ratings.

19 This is one of the oldest projects in the sample and did not have a DEM. However, as pointed out in RE-448-1, many projects have their indicators changed during their execution, whether they have a DEM or not. Therefore, it might be challenging to assign a rating to many projects in the future following this system.

2.18 **During this analysis OVE applied a different rating system that follows more closely what is currently being proposed for private sector operations (Annex III), finding notable differences with the numerical ratings above (see Box 2.1).** Though the results using the different rating scales are clearly correlated, important differences can be seen, especially in the assessment of effectiveness. On the one hand, in a number of cases in which the numerical rating for effectiveness was above 0.5, OVE rated project effectiveness as partially unsatisfactory because the PCR provided insufficient evidence of achievement of the main objectives. On the other hand, in some other projects for which the PCR had sufficient evidence of project achievement, OVE rated effectiveness as satisfactory even if some targets were not achieved. This demonstrates the value of flexibility in assessing project performance.

Box 2.1 Testing an Alternative Rating System on the Pilot Projects

As noted in the report, the current formula-based rating system applied to public sector projects lacks flexibility, which limits opportunities for manipulation but also risks ending up with misleading results. As part of this review, OVE tested an alternative rating system, more aligned with the system used in other MDBs and with the system proposed for private sector operations in IDB. This system uses either a 4- or 6- point scale for each dimension, with half of the ratings being on the positive side (partly, fully, or, in the six-point version, highly satisfactory) and half being on the negative side (partly, fully, or highly unsatisfactory). Effectiveness, for example, is judged using a four-point scale, based on evidence about the project's achievement of its objectives. This approach is more flexible, as the evaluator maintains some discretion to weigh the contribution of each objective to the overall performance of the project and to judge the degree of achievement needed to constitute satisfactory performance. It is not entirely flexible, however, as a strong evidence base is needed to support the rating.

The differences between the two rating systems can be seen in a comparison of two projects in Argentina. AR0202 aimed to facilitate passenger and cargo transportation in priority corridors of Argentina's national road network, while AR-L1008 aimed to strengthen the Argentine Senate by improving its lawmaking processes and the distribution of public information. The current rating system gave AR0202 and AR-L1008 almost the same ratings for effectiveness -- 0.65 and 0.66, respectively. In contrast, OVE's alternative rating system resulted in a satisfactory rating for AR0202 but a partly unsatisfactory rating for AR-L1008. This is because AR0202 made progress in achieving all of its outcomes, despite problems in implementation, while AR-L1008 made no progress toward its most important outcomes. Applying an inflexible formula based on a limited set of outcome indicators could not fully capture the performance of this complex project.

As with effectiveness, OVE rated efficiency of the pilot projects based on evidence provided in the PCRs, focusing, however, on the ex-post economic analysis rather than an ex-ante and ex-post comparison as in the current ratings guidelines. One case, the Nicaraguan-Costa Rican Border Highway Integration Program (NI-L1006), illustrates the difference. The project received a score of 0 under the current rating system, because the ex-post ERR of 13.5% was lower than the ex-ante ERR or 20%. In contrast, OVE's alternative method resulted in a rating of satisfactory, reflecting the fact that 13.5% is a strong ERR.

2.19 **The second problem identified by OVE relates to the PCR template, which lacks space to discuss implementation progress and problems of design.** Although part of this discussion could be included in the section on "Analysis of Vertical Logic,"[20] two-thirds of PCR authors mentioned explicitly that they did not have adequate space to describe issues concerning project design or implementation. PCR authors agreed that problems in implementation and design are some of the main sources of lessons learned, and the omission of such discussion in PCRs limits their learning value.[21]

2.20 **Two other issues are worth considering for greater convergence between public and private sector project self-assessments.** The first concerns compliance with environmental and social safeguards, which is included as part of the sustainability section in the new private sector project evaluation guidelines but is in a separate section in the PCR and does not affect the overall performance score. The second concerns additionality, which is a critical issue in private sector operations. Although no MDB currently considers additionality in project completion reports for public sector operations, it might be worth considering in IDB as the Bank pursues further harmonization between public and private sector evaluation methods.

2.21 **OVE agrees with management's decision to apply the new PCR system only to projects that were approved after 2009, and that therefore have a DEM.** This decision resulted from the pilot, which showed the difficulty of applying the PCR methodology to pre-DEM projects. Given that the analysis of achievements is based on the indicators defined in the results matrix, the quality and monitoring of those indicators are essential to a good PCR, as OVE's review of the pilot PCRs confirmed.[22] In fact, the pilot underscores the importance of strong up-front design for effective ex-post evaluation.

2.22 **In sum, despite the problems pointed out above, the new PCR and its process seem to be much improved from previous practice.** The vast majority of PCR authors in the pilot agreed that this version is

---

20 The 2014 guidelines state: "This analysis intends to document and discuss the evidence supporting the underlying logic in the original project design, i.e. the relationships of cause and effect between the different parts of the results chain (reflected in the results matrix). In particular, the analysis seeks to answer: "What were the main outputs and inputs financed by the project?", "Were these outputs the ones originally identified as necessary for the achievement of project results (outcomes and impacts)?" and "Were the observed results logically linked to these outputs?"

21 PCRs in which this omission is very obvious are ME-L1080 and BR-L1093, where delays are mentioned but no further explanation is provided.

22 The analysis of the pilot shows how difficult it can be to assess the effectiveness of the project when the results matrix does not identify specific, measurable, attributable, realistic, and time-bound (SMART) indicators. Despite progress toward more evaluable projects through the application of the DEM, some recent projects still do not have SMART indicators. In addition, as OVE's recent evaluability assessments have shown, some projects are still approved without a rigorous economic analysis, which is essential for the before-and-after comparison proposed in the new guidelines.

better. It requires more data and evidence to support findings, and it is more integrated with the other tools in the DEF. Therefore, while in the beginning it may be difficult for some project teams to provide sufficient evidence of project performance, over time projects should increasingly benefit from the up-front efforts to define and monitor results indicators and build evaluations into projects, leading to stronger evidence on results.

2.23 **OVE recommends that management continue to adjust the system to enhance its usefulness and accuracy and to promote harmonization with the private sector reporting system.** Adjustments are particularly needed in the rating system, as noted above. In addition, most PCR authors in the pilot stressed the need for division chiefs and managers to give greater attention to PCRs. More time and resources are needed to produce the new PCR with the required quality, and strong institutional support will be needed for the system to function effectively. The benefits should be substantial, however, as IDB and its client countries will be increasingly able to measure and report the results of IDB-financed operations and thereby learn from experience.

III. Private Sector Project Evaluation

A. The New XPSR Guidelines

3.1 **New guidelines are also being designed for self-evaluation of IDB Group (IDBG) private sector operations.**[23] They aim both to solve some existing shortcomings in the current guidelines, which follow ECG-GPS4 (see Box 3.1), and to harmonize evaluation criteria more closely with those for the public sector. OVE is working closely with IDBG's private sector windows on the design of the new guidelines.

3.2 **The new guidelines will move beyond ECG-GPS4 toward an objective-based evaluation methodology, which is expected to raise the focus and the quality of the analysis of IDB effectiveness and to facilitate the identification of a richer and more useful set of lessons.** The core criteria will be the same as those in the public sector PCRs—effectiveness, efficiency, relevance, and sustainability. Additionality, IDB profitability, and work quality will continue to follow formats similar to those currently in use. A broad mapping between the two methodologies, GPS4 and the new guidelines, is presented in Figure 3.1.

---

23 See the proposal for new guidelines, which is currently being piloted with the 2013/2014 batches, in Annex III.

Box 3.1 Methodological Shortcomings of the Private Sector Good Practice Standards (GPS)

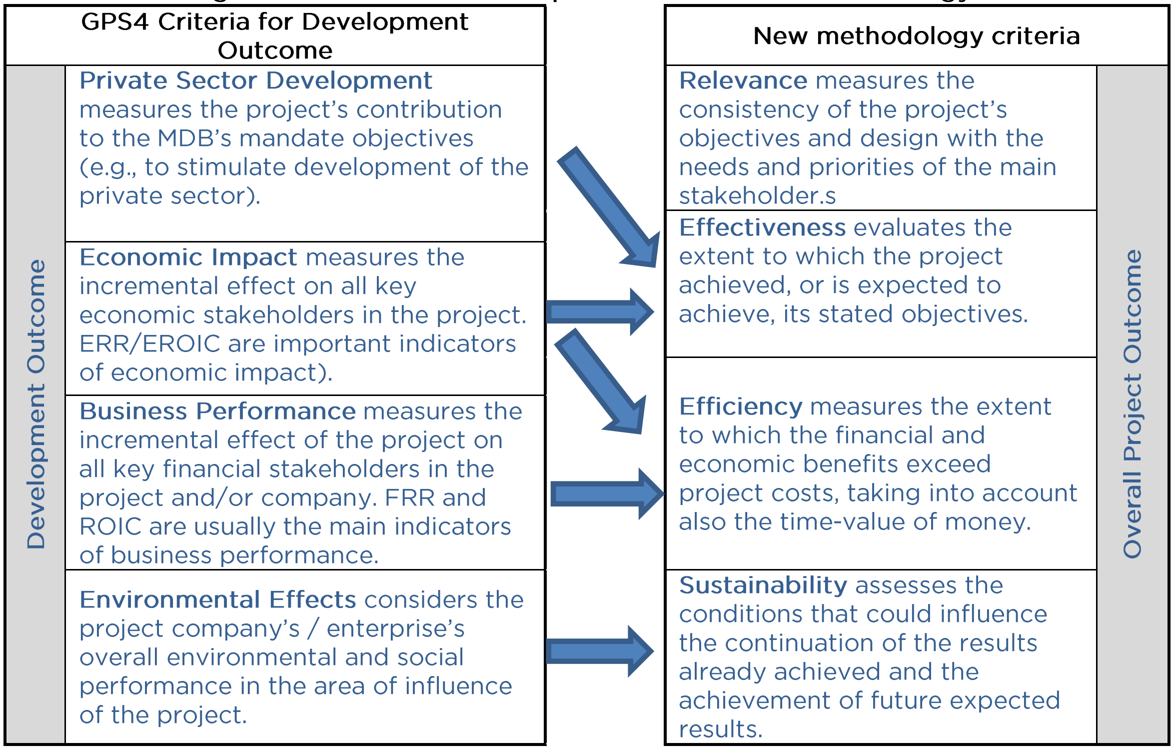
The GPS for private sector operations have come a long way since 1996. They are currently in their fourth edition. GPS4 began to be applied for projects that achieved EOM in 2012, and they attempted to address some of the limitations of GPS3.

The GPS3 assessment of project development outcomes was deficient in three of its four subdimensions: business performance, economic development, and private sector development (PSD). Attribution problems undermined the assessment of the business and economic development contributions, particularly of financial market projects. GPS3 allowed the use of return on invested capital (ROIC) and economic return on invested capital (EROIC) as proxies for financial rate of return (FRR) and economic rate of return (ERR), respectively. However, in many cases these proxies are inadequate since they capture the returns on invested capital of the whole company rather than of the project itself, which can be problematic when projects constitute only a small part of the company's operations.

GPS3 adopted a neutral approach to assess the project's contribution to PSD. If no damage to PSD was found, GPS3 allowed a satisfactory rating even without evidence of positive contribution to PSD. GPS3 also lacked clear indicators for assessing MDB additionality. Although financial additionality can be gauged with comparative evidence from similar operations at the time of project approval, nonfinancial additionality is usually judged through subjective inferences about the MDB's contribution to the project. GPS3 also lacked guidelines linking project evaluability to MDB work quality, allowing projects to receive satisfactory ratings even when they lacked adequate results frameworks or did not track development outcomes during implementation. Finally, GPS3 double-counted the assessment of the Bank's value-added through two separate dimensions, MDB's role and additionality.

GPS4 partially addresses some of these shortcomings. It acknowledges the intrinsic attribution problems of ROIC and EROIC for the assessment of a financial market project's contribution to business and economic development, and it recommends analyzing subportfolios instead, which partially addresses the attribution problem. It also eliminates the MDB's role as a validation dimension, fixing the double-counting issue between MDB's role and additionality. However, GPS4 still falls short in several respects. It does not establish clear evaluation criteria for financial market operations and, more specifically, does not define indicators for the evaluation of financial intermediaries' subportfolios. It does not include evidence-based criteria for rating the project's contribution to PSD. It still lacks assessment criteria for MDB work quality. And it does not have a robust approach to measure project development effectiveness.

Figure 3.1 GPS4 Criteria Compared with the New Methodology



3.3 **The new methodology does not change what is being evaluated but rather how it is assessed and presented.** In the new methodology, effectiveness will be assessed as the extent to which a project achieves or is expected to achieve its stated objectives. The assessment of private sector development and dimensions of economic impact (and, to a lesser extent, business performance) will serve as inputs to the assessment of project effectiveness. Some standard indicators previously used for rating business performance and economic impact – such as FRR or ROIC, and ERR or EROIC – will be used to assess a project's efficiency. The relevance of the project will be rated for the first time, on the basis of the consistency of the project's objectives and design with the needs and priorities of the main stakeholders, including both the Bank Group's and the country's overall strategic development objectives.[24] Sustainability will include social and environmental effects as well as other aspects of the project that influence the durability of results already achieved and the achievement of future expected results. The overall performance rating for the project will be a composite of the scores on the four core criteria.

---

[24] IDBG's key priorities from the IDBG document "A Renewed Vision for the Activities of the IDB Group with the Private Sector," August 2013, are (i) fostering development through the private sector, in particular through the establishment, expansion, and modernization of private enterprises (with a special focus on SMEs); (ii) reducing poverty and social inequalities; (iii) addressing the needs of small and vulnerable countries; (iv) addressing climate change, renewable energy and environmental sustainability; and (v) promoting regional cooperation and integration.

B. OVE's Validation of 2012 XPSRs

3.4 **This section describes the results of OVE's sixth XPSR validation exercise for all seven projects from the Structured and Corporate Finance Department (SCF) that reached EOM in 2012** (the 2013 and 2014 XPSRs are currently piloting the new methodology). The seven loans totaled US$817 million, supporting projects totaling nearly US$6.84 billion. They were approved between 2004 and 2008, and thus none had an evaluability assessment. All but one were in A&B countries – four in Brazil and one each in Colombia, Suriname, and Peru. The loans were made to one debt-equity fund and six large corporations in a variety of sectors (including general manufacturing, energy, infrastructure, and telecommunications). The share of total project cost financed by IDB was 12%. Two loans were prepaid, and one client defaulted, with the XPSR reporting that it had been moved to the Special Asset Unit and was awaiting restructuring. Four loans were given to sponsors that had previously been involved in an operation with the IDB, though none of the specific projects had been supported previously by the IDB. One company later received further support, but in another country.

3.5 **Three projects were accompanied by technical cooperation (TC), which helped bring nonfinancial added value to them.** The IDB's Sustainable Energy and Climate Change Initiative supported a client's efforts to improve its environmental footprint through an inventory and reduction of greenhouse gases and expansion of efficient technologies; however, it was also being supported through the public sector window, so it is unclear how much this TC can be attributed to this loan in particular. The Sustainable Energy and Climate Change Initiative also provided a TC for an energy efficiency assessment of a second project. For a third project, IDB delivered three TC grants amounting to US$2.89 million to strengthen the Peruvian Government's regulatory and social agencies. The TCs developed an energy matrix of the hydrocarbon sector and a strategic plan for sustainable energy and biofuels development, and it supported the economic and social development of native communities.

3.6 **SCF coordinated with the Bank's public sector on two projects, one in energy and one in water and sanitation.** There is little evidence of SCF coordination with other private sector windows, although SCF cofinanced one loan with IIC. All projects included the standard participation of the Environmental Safeguards Unit (ESG), Legal Department (LEG) and the Portfolio Management Unit (PMU).

3.7 **In line with the GPS4 methodology, projects were rated on four performance dimensions: Project Development Outcome, SCF Investment Profitability, SCF Additionality, and SCF Work Quality** (Table 3.2).[25] Each of the performance dimensions was assigned a rating[26] based on a standard four-point scale: Excellent (E), Satisfactory (S), Partly Unsatisfactory (PU), and Unsatisfactory (U).[27] The rating for each dimension was derived from ratings on individual subcategories, or performance areas, as shown in Table 3.1.

3.8 **The development outcomes of the projects in this 2012 batch were generally positive** (Table 3.2).

- First, the contribution to company business performance was rated positive in five of the seven projects. To assess the incremental effect of the project on the company, projects tracked financial performance through several indicators: estimates of FRRs/ROIC, achievements of other business objectives (many of which were not quantifiable), and analysis of financial statements of companies. Business objectives usually related to company production, sales, profitability, capital structures, or corporate governance. Two projects earned an excellent rating, as they provided solid financial returns and exceeded appraisal projections. However, two projects failed to contribute to the clients' business success, as they were not able to achieve the financial returns expected at approval, and in one (rated unsatisfactory) the company failed to meet obligations to creditors.

---

25 Development Outcome aims at assessing the project's contributions to the host country's overall development, taking into account the four underlying areas and the project's sustainability. Investment Outcome reflects SCF's financial returns from the project. Work Quality focuses on SCF's role and operational effectiveness. Additionality assesses whether the project added further value by catalyzing private sector investments, providing terms not available in the market, or improving the project's design or functioning.

26 As in the previous validation exercises, OVE and SCF established an interactive process that increased the number of agreements. This time, given that it is a period of transition to the new guidelines, OVE used the first instance of interactions to ask for more evidence when needed. After this interaction, a few disagreements remained, resulting in 13 downgrades (10 in Development Outcome, 2 in Additionality and 1 in Investment Outcome). Four upgrades were made in the subcategories, i.e. in Business Performance and Economic Impact, but did not affect the overall categories.

27 See Annexes II and III for a detailed description of ratings criteria and benchmarks.

- Second, project contribution to economic development[28] was rated positive for the majority of operations, and two of the seven were rated excellent. One of these, had a higher than expected economic and social return and was able to increase service reach and quality, especially in poor and underserved areas. In contrast, another project failed to contribute positively to economic development. Though the project led to an increase in the number of new customers, the company failed to meet the minimum requirements for a positive financial performance and therefore generated net economic costs for stakeholders.

Table 3.1 Performance Dimensions and Standard Ratings

Performance dimension	Performance areas
Project Development Outcome	Contribution to Business Performance
	Contribution to Economic Impact
	Environmental and Social Effects
	Contribution to Private Sector Development
SCF Investment Profitability	Loan Performance
	Equity Performance
SCF Additionality	Financial and Nonfinancial Additionality
SCF Work Quality	Screening, Appraisal, and Structuring
	Monitoring and Supervision Quality

Note: Table extracted from the 5th XPSR Validation Report, RE-332-8.

28 The projects were intended to achieve a variety of economic development benefits, including increased employment, strengthened supply chains, improvements in infrastructure, promotion of green energy, and generation of foreign exchange. Projects tracked the contribution to economic development through either a quantitative method or a qualitative stakeholder analysis. Estimates of ERR/EROIC often poorly quantified the projects' economic and social effects on population—in most cases only adding taxes paid to financial returns. All projects also included a stakeholder analysis and tracked the achievement of other development objectives. However, evidence and documented results were poor in most cases.

Table 3.2 Performance Distribution by Number of Projects

OVE ratings/General Performance (%)	Excellent	Satisfactory	Partly unsatisfactory	Unsatisfactory
PROJECT DEVELOPMENT OUTCOME	3	3	1	0
Business Performance	2	3	1	1
Econonmic Impact	2	4	1	0
Environmental Effects	3	4	0	0
Private Sector Development	2	5	0	0
SCF's INVESTMENT PROFITABILITY	1	5	0	1
SCF's ADDITIONALITY	2	5	0	0
SCF's WORK QUALITY	1	5	1	0
Screening, Appraisal, Structuring	1	5	0	1
Supervision and Administration	2	5	0	0

- Third, all projects performed well on environmental effects. In general, projects rated satisfactory were in compliance with SCF's environmental and social guidelines as well as with domestic regulations.
- Finally, all projects achieved a positive rating on PSD. The GPS4 has a low bar, recommending a satisfactory rating in the absence of verifiable damage to the private sector. Therefore, the assessment of the PSD dimension for both financial and nonfinancial operations tends to be biased toward positive ratings. A few projects in this exercise did show sufficient evidence of positive contribution to PSD. For instance, through one project, the Bank helped establish a type of fund that did not exist before.

3.9 **Scores on criteria other than project outcomes were also relatively positive in this batch.** First, all projects but one received positive ratings on profitability.[29] Two loans were prepaid and one client defaulted on its debt in May 2012, which resulted in an unsatisfactory rating for IDB profitability. At the time of the evaluation, this project was at Special Asset Unit awaiting restructuring.

3.10 **All projects were rated positively on additionality.**[30] All provided arguments to justify IDB's financial additionality, mostly mobilization of B-lenders or provision of longer tenors than those available in the

---

29 This rating is expected to provide comparative information on the project's contribution to IDB profit. In practice it is assessed by comparing the fees and interest collected during the project's life cycle with the amounts expected to be received at approval. The project's loan credit risk classification is also taken into account in the rating process.

30 It is important to note that under the new methodology projects will need to demonstrate at least one type of additionality, but not both. With this in mind, OVE was more flexible in the its rating, provided that the case for financial additionality was made.

market at the time of project approval. In most cases the conditions offered by SCF were not available in the market (particularly during the financial crisis), implying that the project would otherwise fail to go ahead with financing on "appropriate terms and/or with undue delays." Little evidence was available to justify nonfinancial additionality, though the projects followed GPS4 standards by strengthening "the company or [financial intermediary] in its design, business, developmental, transition, social or environmental terms".[31]

3.11 **Finally, all projects received a positive rating on work quality, except for one project that received an unsatisfactory rating on Screening, Appraisal and Structuring.** That project faced problems in identifying important risks and mitigation measures affecting project implementation and lack of evidence that these events could not have been anticipated.

3.12 **The new guidelines will for the first time require a results framework with measurable indicators to gauge project results, aided by the recent introduction of the DEM for non-sovereign-guaranteed (NSG) operations.** This 2012 batch of projects generally did not have such results frameworks. In some cases the projects defined beneficiaries but did not quantify expected benefits, hindering the ex-post calculation of economic development results. These shortcomings were not captured by the work quality rating, since the GPS4 does not require such a framework. It is important to keep in mind that the results framework for a good ex-post project assessment has only been introduced in SCF projects in 2014. Therefore, there will be a long transition period in which projects will have to be considered with caution.

3.13 **Though the XPSRs' discussion of lessons learned is still relatively weak, it has generally improved, particularly regarding development effectiveness, as compared to previous validation exercises.** Three XPSRs discussed lessons on IDB profitability, two lessons on additionality, and six lessons on IDB's work quality.[32] Lessons related to work quality tend to be better structured and more useful for future operations,[33] and SCF needs to ensure that these lessons are shared with investment officers and incorporated into SCF operations. In addition, SCF has addressed this issue by improving projects' evaluability, and it will likely be evident in future XPSRs. As mentioned above, none of the projects in this batch had a DEM at entry.

---

31 In many cases SCF was required to clarify some questions and provide additional evidence, to allow OVE to validate the nonfinancial additionality rating. In one project, OVE did not accept the nonfinancial additionality.

32 Most lessons tend to be more of a conclusion of the experience with the project rather than an actual lesson learned. Although interesting comments can be found in the lessons learned section, they usually miss a reflection of how that particular experience can be translated into a lesson for SCF.

IV. Summary and Recommendations

4.1 **This report reviews recent developments in project-level results reporting for both the public and the private sector windows of IDB.** On the public sector side, it assesses the PCR pilot exercise and provides recommendations to assist in the implementation of the new PCR system. On the private sector side, it describes efforts to design a new format for results reporting and completes the cycle of validations for private sector XPSRs under the prior GPS4 methodology.

4.2 **The first finding from this report is that the use of different systems for public and private sector operations can often generate confusion.** Up to now, public and private operations have been assessed according to different criteria and using different rating systems. Therefore, it is virtually impossible, under the current systems, to compare the effectiveness of public and private sector operations, which would be an interesting exercise for an organization like the IDB.

4.3 **OVE's review of the PCR pilot phase found that the overall quality of the reports was much higher than that of earlier PCRs, but there are still shortcomings in the PCR system.** The analyses produced under the PCR pilot were generally supported by data, and a clear effort was made to better document results achieved. Most problems identified by OVE arose from shortcomings in the guidelines, most notably an inflexible rating system and inadequate space to discuss issues of implementation or ex-post problems with project design. Low evaluability also affected the quality of PCRs for older projects, which has led Management to apply the new PCR system only to projects approved after 2009.

4.4 **A main drawback of the current PCR rating system is that it diverges from the standard approach used by the IDB group private sector windows, hindering their comparability.** Other MDBs and IDB's private sector windows use a discrete 4- or 6-point rating system, which the PCR guidelines use a continuous numerical rating for effectiveness and overall project performance, and a 5-scale for other core criteria.

4.5 **For the XPSRs reviewed, the ratings (following GPS4) were generally positive, but the quality of the XPSRs could be further improved.** Most XPSRs did not document well the effectiveness of projects, partly because of limitations in GPS4. They often did not discuss the results chain, hindering the connection between claimed outcomes and project activities. Strengthening discussion of the causal chain and the evidence on effectiveness, in line with the new guidelines, should improve the quality of the XPSRs and strengthen lessons learned.

4.6 **The key change in the new XPSR methodology, compared to the ECG-GPS, is to focus on the projects' objectives as the basis for the evaluation.** Projects will be evaluated against the results (impacts, outcomes, outputs) they were intended to achieve. NSG operations can focus on private sector development objectives, but can also aim to contribute more broadly to the host country's development. In addition, the XPSRs will more clearly and directly assess the relevance of the operation ex-post and the attribution of observed outcomes to IDBG activities.

4.7 **The Bank is making excellent progress toward a robust and unified project results measurement and reporting system that will facilitate accurate and comparable results reporting across all parts of the IDB Group. Looking forward, OVE has the following two recommendations for management to facilitate that progress**:

1. **Revise the PCR guidelines to further harmonize them with those for the private sector and to address some shortcomings identified in this report. In particular:**

    a) **Revise the rating system.** OVE recommends that management work in partnership with the private sector windows to harmonize their rating systems, using a 4-point scale for all core criteria and a 6-point scale to assess the overall performance of the projects. The current system of continuous numerical ratings could be used as input but would not constitute the final ratings reported for projects. In validating management's project ratings, OVE would take into consideration any numerical ratings but would also use discretion to arrive at what it considers to be the most sensible final project ratings.

    b) **Include space for discussion of implementation progress and an ex-post analysis of project design.** One of the main weaknesses of the new guidelines is the omission of discussion on implementation progress and project design, which also limits the discussion of lessons learned.

    c) **Include compliance with environmental and social safeguards as a factor in sustainability.** As in the private sector guidelines, the PCRs should consider compliance with safeguards as a dimension of sustainability rather than have it as a separate non-core criteria as in the current PCR guidelines.

2. **Continue the process of implementing the new XPSR guidelines to improve SCF's and OMJ's ability to report on results, and to further harmonize with the other IDB windows.** As the Bank moves towards the merger of all private sector windows into IIC, it is more important than ever that the private sector windows harmonize their systems to report on results. The new methodology has a stronger focus on development results and is more harmonized with the public sector methodology than the previous syste,, and it will enhance the rigor with which information is presented. It will further reinforce the need for projects to be evaluable at approval.