Project Completion Report Analysis: Factors Behind Project Success and Effectiveness

Office of Strategic Planning
and Development
Effectiveness

Carola Álvarez
Leonardo Corral
Ana Cuesta
José Martínez
César Montiel
Consuelo Yepez

TECHNICAL NOTE Nº
IDB-TN-02135

March 2021

Project Completion Report Analysis: Factors Behind Project Success and Effectiveness

Carola Álvarez
Leonardo Corral
Ana Cuesta
José Martínez
César Montiel
Consuelo Yépez

Inter-American Development Bank
Office of Strategic Planning and Development Effectiveness

March 2021

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Project completion report analysis: factors behind project success and effectiveness / Carola Álvarez, Leonardo Corral, Ana Cuesta, José Martínez, César Montiel, Consuelo Yépez.
p. cm. — (IDB Technical Note ; 2135)
1. Economic development projects-Management. 2. Economic development projects-Evaluation. 3. Project management. I. Álvarez, Carola. II. Corral, Leonardo. III. Cuesta, Ana. IV. Martínez, José. V. Montiel, César. VI. Yépez, Consuelo. VII. Inter-American Development Bank. Office of Strategic Planning and Development Effectiveness. VIII. Series.
IDB-TN-2135

http://www.iadb.org

Leonardo Corral | leonardoc@iadb.org

Project Completion Report Analysis:

Factors Behind Project Success and Effectiveness[§]

Carola Álvarez[a], Leonardo Corral[a*], Ana Cuesta[a], José Martínez[a], César Montiel[a], Consuelo Yépez[a]

[a] Strategy Development Division, Office of Strategic Planning and Development Effectiveness, Inter-American Development Bank

Abstract

The goal of development assistance is to deliver measurable results. At the Inter-American Development Bank (IDB), the Development Effectiveness Framework (DEF) was adopted to enhance the likelihood this goal is met. The key objective of this analysis is to enhance our understanding of the role the set of tools and processes adopted at the IDB, from project design to completion, play in the assessment of a project's success and effectiveness. Specifically, we are interested in teasing out the role different dimensions of project design, assessed through the Development Effectiveness Matrix (DEM), and execution performance indicators, as per the Project Monitoring Report (PMR), have on the delivery of effective and successful projects measured through the Project Completion Report (PCR) methodology implemented at the IDB. We also explore the validation process of PCR's and delve into the effectiveness analysis at the level of specific objectives and result indicators, identifying key pitfalls associated with objectives not being met. Overall, our findings provide strong validation for the DEF and its tools. We find robust evidence for the importance of the quality at entry assessment conducted through the DEM. In particular, the quality of the results matrix at approval is a strong determinant of project success and effectiveness at closure, and the quality of the Evaluation Plan is strongly associated with the effectiveness of a project. In terms of execution performance, our analysis finds that, on average, projects that execute a lower share of their approved loan amount, are put on Alert, or are classified as a Problem, in its first three years of execution; and projects that experience a higher share of their outputs discontinued, with respect to their first results matrix, are most likely to be ineffective in achieving their objectives and will likely be rated as unsuccessful.

JEL classifications: O1; O12; O19; O22

Keywords: Development Effectiveness; Quality-at-entry; Project Performance

[§] We are grateful to Shakirah Cossens, Mario Rodriguez and Juan Israel Zeledon for supporting data access and management and PMR specific knowledge. Oscar Mitnik and Sandra Bartels provided helpful comments.
[*] Corresponding author: Leonardo Corral. E-mail address: leonardoc@iadb.org

Introduction

The goal of development assistance is to deliver results. As such, during the last decades, the concept of development effectiveness has progressively gained strategic relevance within development institutions (United Nations Development Program [UNDP], 2001; World Bank, 2005; Crespo et al., 2013; International Fund for Agricultural Development [IFAD], 2014b; Asian Development Bank, 2015; Carnairo and Garbero, 2017; Corral and McCarthy, 2018).

This movement has materialized into an important agenda of the international community, of which a major milestone is the 2005 Paris Declaration on Aid Effectiveness, endorsed by the Inter-American Development Bank (IDB). At the same time, the harmonization efforts carried out by the Multilateral Development Banks (MDB) have been noteworthy. Through the Evaluation Cooperation Group (ECG), MDBs have formulated and agreed upon common standards and good practices for evaluation. Finally, the MDBs have also set in motion several initiatives to support management for results, such as the Common Performance Assessment System (COMPAS), which establishes a common framework for MDBs to report results.

At the IDB, the Development Effectiveness Framework (DEF) was adopted in 2008 to align with these efforts and enhance the likelihood that the projects it supports deliver measurable results. The instruments of the DEF (Development Effectiveness Matrix (DEM), Project Monitoring Report (PMR) and Project Completion Report (PCR)) were designed with the hope that each would contribute to assure that projects reach their development objectives. In a sense, one can imagine a production function for delivering projects that attain their results, where the key factors are the quality at entry (captured by the DEM); the performance during execution (captured by the PMR indicators); and other context specific factors that might affect project success. The end result is captured by the PCR, where final reporting and assessment of a project's success is prepared by IDB's management (MGT) and externally validated by the Office of Evaluation and Oversight (OVE), which reports directly to the Board of Directors of the IDB.

The key objective of this analysis is to enhance our understanding of the role the set of tools and processes adopted at the IDB, from project design to completion, play in the assessment of a project's success using the last two cycles of PCR validations (i.e. for projects that reached operational closure (CO) in 2017 and 2018). Specifically, we are interested in teasing out the role different dimensions of project design, assessed through the DEM, and execution performance, as per PMR indicators, have on the delivery of effective and successful projects. The reader should be aware that there are numerous unobserved, or difficult to quantify, factors

that can affect a project's effectiveness, ranging from the country to the sector, to the executing agency and its personnel characteristics that we are not able to control for due to data limitations.

From here forward, we will use PCR17 and PCR18 to identify PCRs for projects that reached CO in 2017 and 2018, respectively. Unless otherwise noted, the analysis presented in this report is based on data from 97 PCRs, of which 40 are PCR17 and 57 PCR18. To make the observations across years as comparable as possible, only PCRs that were prepared or validated under similar guidelines are used in this investigation.

This analysis has been spurred by the lackluster performance of PCR18 in relation to PCR17. As such, the first line of inquiry will aim to better understand what is behind this drop in success and effectiveness rate[1]. Using OVEs ratings, 67.5% of PCR17 were rated partly successful or above compared to 50.9% for PCR18. This represents a 16.6% drop. Similarly, 35% were rated effective by OVE for PCR17 compared to 26.3% for PCR18, an 8.7% drop. We also assess differences in the validation process from PCR17 and PCR18, looking at the divergence between Management and OVE's ratings for the project overall and its core criteria, and assess the implications the divergence of ratings on each of the core criteria has through the use of simulations. Using simulations, we also attempt to shed light on the implications of the weighting of criteria and the scoring prescribed by the PCR Guidelines.

Given the importance the Effectiveness rating has on overall project success, in a subsequent phase of this analysis, we delve into factors associated with the effectiveness rating. Specifically, we review the 152 specific objectives (SO) of 62 PCR18 validated by OVE, systematically identifying common issues and coding these for our analysis. We attempt to differentiate between factors that affected project effectiveness related to design and execution, and factors related to methodological discrepancies between OVE and Management. We also present an analysis at the level of the 515 result indicators of PCR18 and we assess the impact the Effectiveness rating formula has on the overall Effectiveness of PCR18 with simulations.

[1] For purposes of the analysis carried out throughout we consider a project to be successful if its overall PCR rating is Partly Successful or above (=>Partly Successful). This is primarily done on account of the limitation on the number of PCRs available for the analysis and thus the lack of statistical power to discern effects at a lower level of aggregation. A similarly, approach is taken with regards to specific core criteria. For instance, a project is considered to have been effective in achieving its development objectives if its rating for these criteria is Satisfactory or above (=>Satisfactory).

But even if we consider PCR18 an anomaly, the overall effectiveness of projects is well below what one would hope for[2]. Thus, in a final line of inquiry we will attempt to assess the role the DEF, and other factors, play in determining a project's effectiveness and overall success as captured by PCR ratings. We will attempt to explain, for both MGT and OVE ratings, what determines (or is associated with) whether a project is rated successful or not and effective or not. The explanatory variables include the quality of the project at entry (DEM), its performance in execution (PMR), and other characteristics such as the sector a project belongs to. An anticipated limitation of this analysis lies with the relatively few observations available. Thus, it is expected that this type of analysis can be routinely reproduced in future, as more projects achieve CO and are assessed for their effectiveness and success.

Our analysis finds no evidence that PCR18 were less evaluable at entry than PCR17. We do find some evidence that PCR18 execution performance was more lackluster than PCR17, however this difference was not robust enough to be statistically meaningful.

In terms of the validation process, comparing PCR17 to PCR18, two aspects are noteworthy. First, there was greater divergence between MGT's rating and OVE's for PCR18. Second OVE's ratings for PCR18 show greater mobility, in the sense that the disagreement in ratings for specific criteria between MGTs and OVE is greater than for PCR17. We also find some evidence suggestive of OVE's ratings being more compartmentalized for PCR17 as opposed to PCR18. That is, there appears to be greater correlation among ratings from one criterion to the next for PCR18, resulting in a negative rating for one criterion influencing the rating of the remaining criteria.

With regards to the analysis of OVE's assessment of the achievement of specific objectives, the main deficiencies associated with close to 64% of specific objectives being rated unfavorably can be categorized as 1) issues related to execution, which explain 53.1% of unfavorable rating of specific objectives; 2) issues related to project design, which explain 25% of unfavorable rating of specific objectives; 3) methodological discrepancies between OVE and Management, which explain 18.8% of unfavorable rating of specific objectives; and 4) other reasons, which explain 3.1% of cases. Execution issues included cancelation of outputs and monitoring pitfalls that did not allow for reporting on the original indicators. Issues in the design phase that affected achievement of specific objectives included: i) unrealistic targets, ii) weak results matrix, and iii)

[2] For instance, the IDBs Corporate Results Framework (CRF) 2016-2019 established a target of 80% of operations with satisfactory development results at completion. The CRF for 2020-2023 establishes a target of 70%.

specific objectives with no metrics to value achievement. Finally, methodological discrepancies between OVE and Management included: OVE not accepting additional indicators or the attribution analysis presented; OVE altering the results matrix by including or excluding indicators; and OVE adding an specific objective that was not originally included for the validation.

We also find that despite the low achievement at the level of specific objectives, 48% of result indicators for PCR18 were fully achieved. However, nearly 10% of result indicators had no information of achievement, indicating serious monitoring failings. We do find that relatively small changes to the Effectiveness rating formula can increase the overall effectiveness of PCR18 projects, perhaps better aligning achievement at the level of specific objectives and achievement of result indicators.

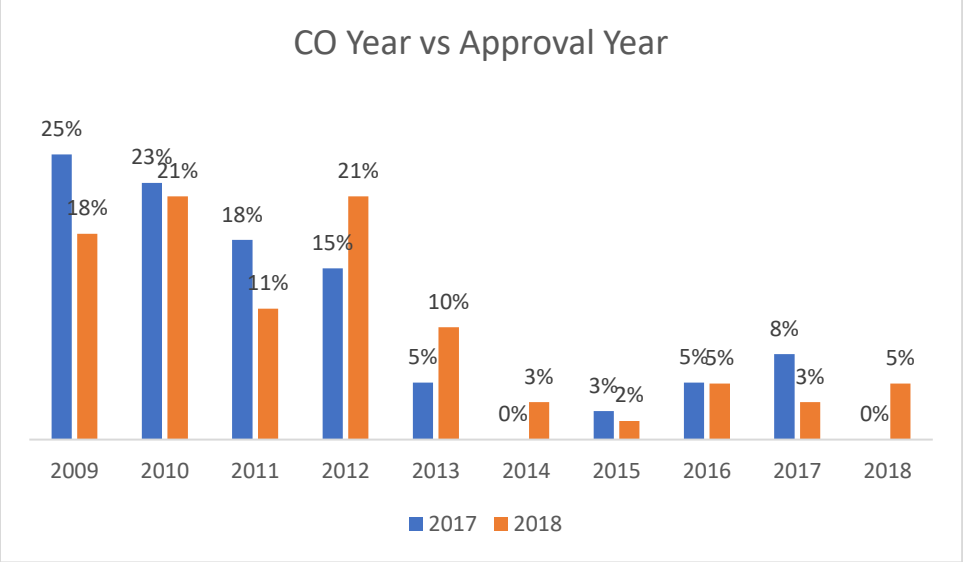
Overall, our findings provide strong validation for the DEF and its tools. We find robust evidence for the importance of the quality at entry assessment conducted through the DEM. In particular, the quality of the results matrix at approval is a strong determinant of project success and effectiveness at closure, and the quality of the Evaluation Plan is strongly associated with the effectiveness of a project. In terms of execution performance, our analysis finds that, on average, projects that execute a lower share of their approved loan amount; are put on Alert, or are classified as a Problem, in its first three years of execution; and experience a higher share of their outputs discontinued, with respect to their first results matrix, are most likely to be ineffective in achieving their objectives and will likely be rated as unsuccessful.

This report is organized as follows. In the next section we attempt to enhance our understanding of factors behind the observed drop in PCR18 with regards to PCR17. Next, we explore the main deficiencies found in PCR18 with regards to their achievement of specific objectives. We then assess the role the DEF, and other factors, play in determining a project's effectiveness and overall success as captured by PCR ratings. Lastly, we present some conclusions and recommendations.

Descriptive statistics for PCR17 and PCR18

Figure 1 presents the year the project achieved CO by approval year. As expected, PCR18 is comprised of relatively newer projects – half were approved from 2012 on – whereas 66% of PCR17 were approved before 2012.

Figure 1. CO Year by Approval Year



The sectoral composition by year of CO is presented in Figure 2. Marked increases are appreciable for SPH and reductions for FMM and RND across CO year.

Figure 2. Sectoral Composition by Year of CO

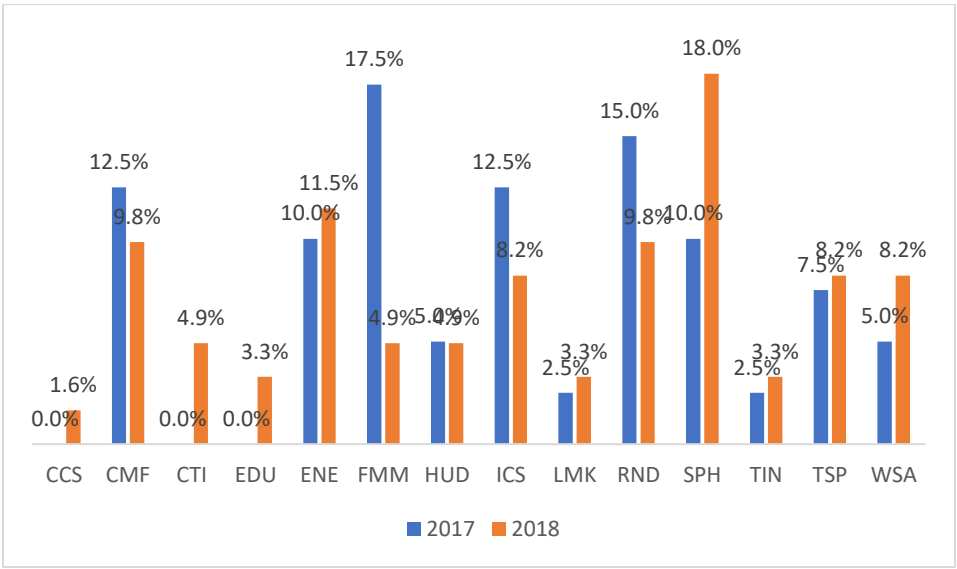


Table 1 presents descriptive statistics by CO year for sectors and country departments and statistical tests of the difference across years[3]. This analysis is based on 40 PCRs with CO in 2017 and 57 with CO in 2018, which were prepared and validated under the same PCR directives (PCR 2018 Guidelines). For purposes of this analysis, 4 observations with CO in 2018 have been dropped – 2 CCLIPs and 2 grants—due to missing values. In addition, there were 4 instances where MGT ratings were not in line with the PCR score. For example, the PCR for CH-L1136 was self-rated by the PCR team as Highly Successful, although according to its PCR score and the PCR Overall Rating classification, it should have been rated as Successful. In these cases, we revert to the classification as per the PCR Guidelines.

In terms of the distribution across Vice-Presidency of Sectors (VPS) and Vice-Presidency of Countries (VPC) Departments, the only statistical difference is an increase participation of SCL in PCR18, driven by SPH, and a reduction in the participation of CAN. A similar proportion of PBLs, makes up the portfolio across both years.

PCR Ratings

Table 2 presents PCR ratings by Management and OVE by CO year. In terms of MGT PCR ratings, there is a marked reduction (14%) in the ratio of PCRs rated as Partly Successful from PCR17 to PCR18. Most of this reduction appears to be reflected in an increase in the percentage of PCRs rated Partly Unsuccessful and Unsuccessful. MGT self-rating for project success (>= Partly Successful) decreased from 80% for PCR17 to 68% for PCR18, close to a 12% reduction.

This trend is also present in OVE PCR ratings, but proportionally more pronounced. The largest drop is in the Successful category, which went from 27% for PCR17 to 19% for PCR18. OVEs rating for project success (>= Partly Successful) decreased from 67% for PCR17 to 51% for PCR18, almost a 17% change

Turning to the core criteria, apart from Relevance, MGT rated PCR18 Effectiveness, Efficiency and Sustainability slightly more satisfactory (>=Satisfactory) than PCR17. However, this measure masks some important variability within criteria, which reveals itself when looking at MGT relevance, effectiveness, and sustainability scores. MGTs relevance score decreased from 3.675 to 3.421, a significant reduction. MGTs effectiveness score also decreased from 2.4 to

[3] Throughout this document, robust Standard errors are presented in parenthesis. ^, *, **, *** depict statistically significant at 0.20, 0.10, 0.05, and 0.01, respectively. The use of statistical significance at 0.20 is normally not in current practice, however, for the purposes of this analysis, we include.

2.2, barely above the threshold for Partly Unsatisfactory (PU) Overall, this decrease in relevance and effectiveness score is driving a reduction in MGT's PCR score from 2.92 for PCR17 to 2.79 for PCR18.

Again, this trend is also evident in OVEs scoring, but magnified. Significant drops in OVEs Relevance and Effectiveness Score from PCR17 to PCR18 are observed. This is reflected in OVEs PCR Score reduction of 2.7 for PCR17 to 2.49 for PCR17. Annex A presents a disaggregated analysis by VPS Departments.

DEM Dimensions

Given this observed decline in the success and effectiveness of projects from PCR17 to PCR18, we turn now to project characteristics associated with quality at entry, and performance during execution. Table 3 presents descriptive statistics by CO year of DEM scores by dimensions and other project characteristics. We find no evidence that PCR18 were less evaluable at entry than PCR17. If anything, except for Results Matrix Quality and Monitoring, which exhibit very small declines, most other dimensions of quality at entry suggest that PCR18 were slightly more evaluable, although no difference is statistically significant. In terms of other project characteristics, approved amount for PCR18 was smaller, PCR18 had on average more components, a slightly smaller number of outcomes per component, but more outputs per component. None of these differences was statistically significant.

PMR Indicators

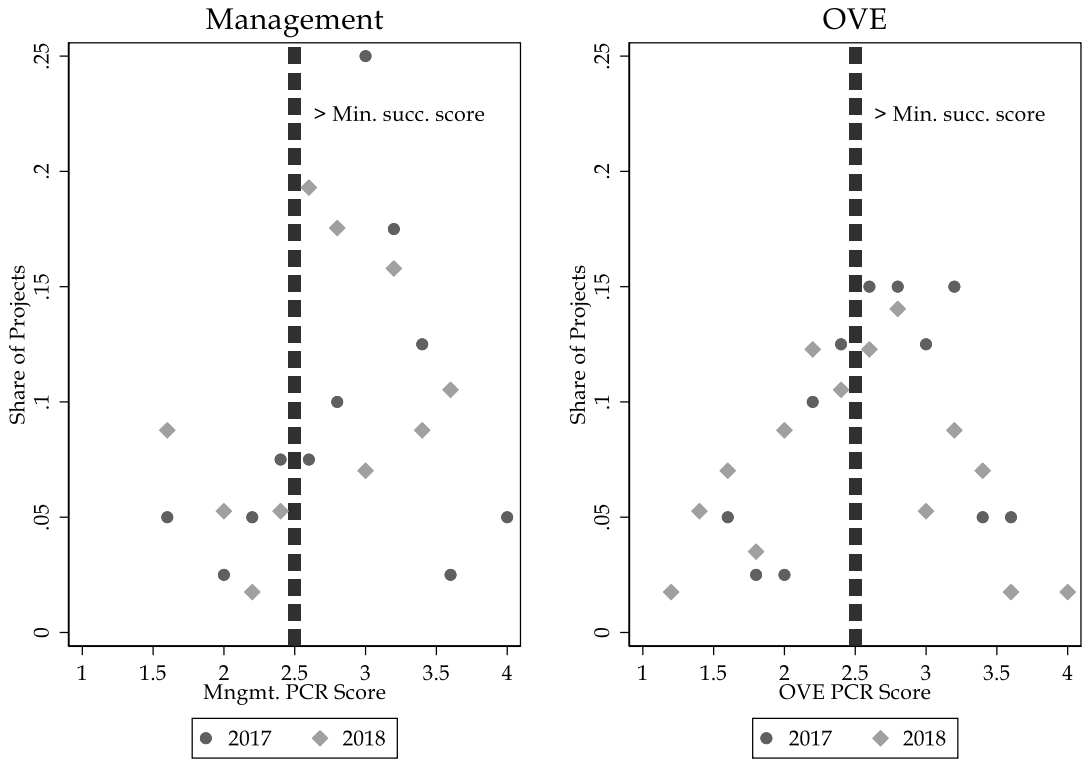
In terms of performance during execution (see Table 4), the opposite seems to be true. PCR18 had an overall lower disbursement rate, a higher proportion (about 5%) disbursed less than 60% of the original approved amount, took longer to reach eligibility after approval, and experienced over 13 months of overrun compared to 10.6 months for PCR17. PCR18 also had a higher share of outputs not achieved. However, none of these differences were statistically significant.

The PCR Score

We turn now to explore potential differences in the PCR score between PCR17 and PCR18. For this we exploit the PCR score variable in a regression discontinuity context. The OVE PCR Score for PCR18 is 2.491, which falls just short of the threshold needed of 2.5 for the whole portfolio of PCR18 to be on average rated as >= Partly Successful (abstracting for now from the

additional rules for the overall rating calculation)[4]. Figure 3 presents the PCR Score for OVE and MGT for PCR17 and PCR18.

Figure 3: Presence of clustering of projects around minimum successful PCR score



Note: The PCR score goes from a minimum of 1 to a maximum of 4. The minimum score to get a partly successful grade or more is 2.5 (=(1+4)/2). However, be aware that there some additional conditions to pass (see footnote 3).

The first thing to notice from the Figure is that whereas for MGT there appears to be a clear bunching of projects just above 2.5 for PCR18[5], this is not the case for OVE's PCR score. Thus, in Figure 4 and 5 we present project characteristics related to quality at entry and execution performance around the threshold of 2.5. We find no significant difference in characteristics for projects that are just above or just below the threshold, with the exception of the share of outputs not achieved form the last results matrix, which is significantly higher for projects just

[4] According to 2018 PCR Guidelines, the PCR Overall Rating is a function of the PCR Score, where a score of >=2.5 is Partly Successful, and additional rules: (i) With unsatisfactory rating for relevance, effectiveness, sustainability: highest achievable rating is partly unsuccessful; (ii) For a highly successful rating, no core-criteria rating can be less than satisfactory; (iii) With a partly unsatisfactory rating for relevance, effectiveness or sustainability, highest achievable overall rating is partly successful. These additional rules can be binding. On account of rule (i) 10.5% of projects where downgraded to PU for PCR18, whereas none was for PCR17. Because of rule (iii), 30% of PCR17 were downgraded to PS, whereas 7% were for PCR18.

[5] The fact that no bunching was apparent in MGT PCR17 and clear clustering above 2.5 is apparent for PCR18, would suggest that there has been some "teaching to test" effect of the PCR workshops held on 2019.

below the threshold. Notwithstanding this last variable, this analysis suggests that PCRs that fall just below or above 2.5 are not different in any significant way in terms of the characteristics assessed. More provocatively, this result seems to suggest that what is perceived as a drop in the rating of success for PCR18 is partly the result of the artifice of the rating scheme.

Figure 4: Average characteristics around OVE PCR score threshold – Panel A (Pooled PCR17 & PCR18)

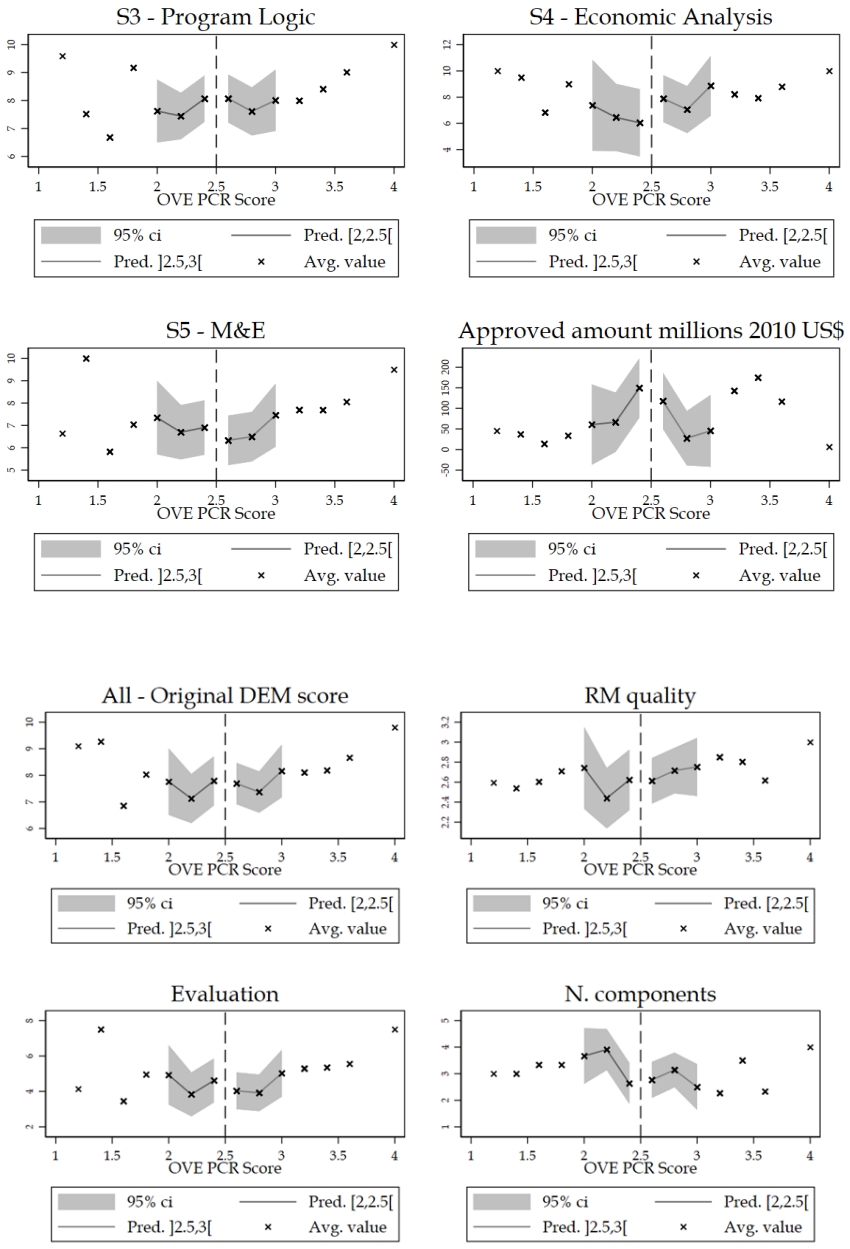
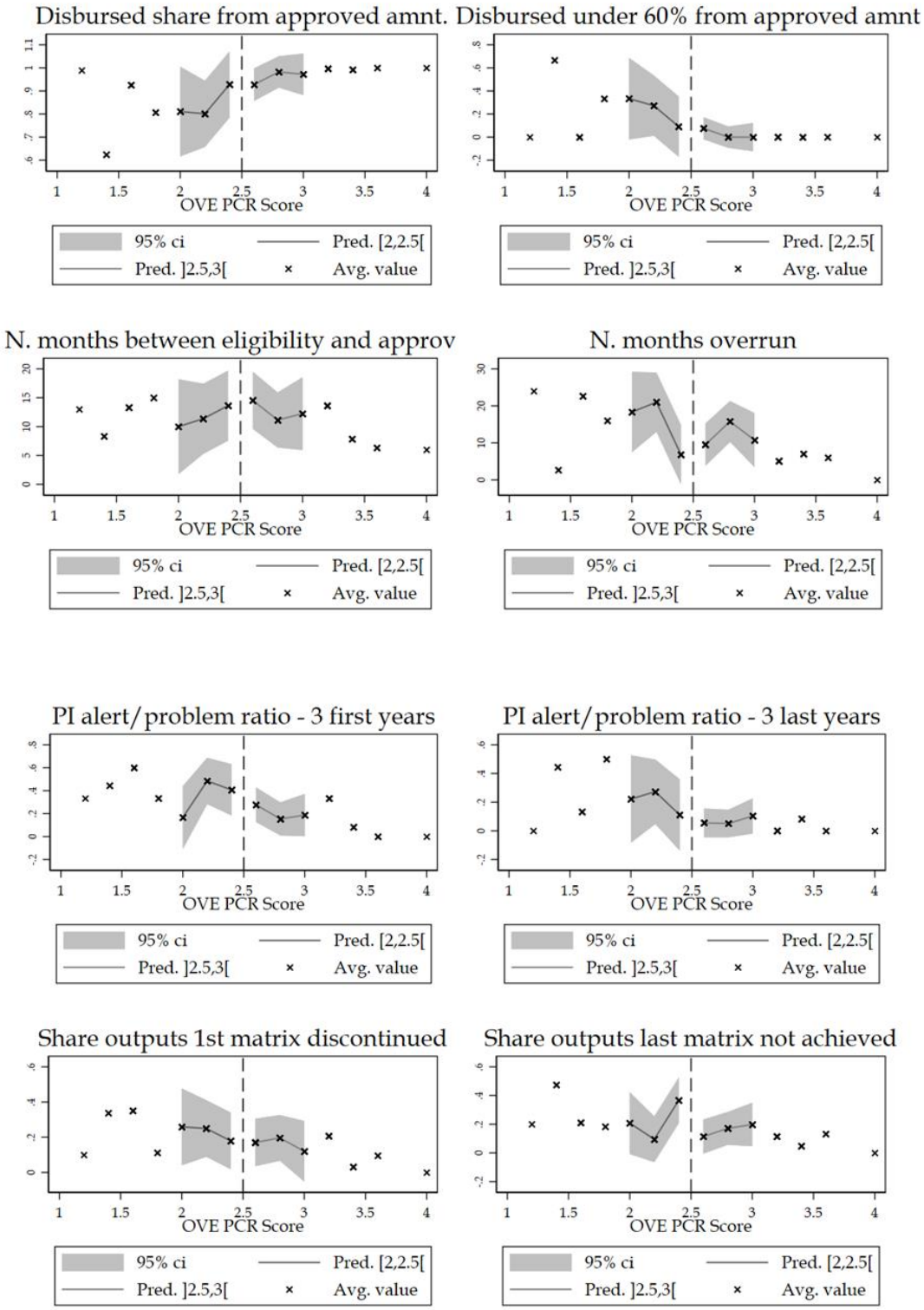


Figure 5: Average characteristics around OVE PCR score threshold –

Panel B (Pooled CO 2017 & CO 2018)



Simulation Analysis

To better understand the impact on project success of changes to MGT's ratings due to OVE's validation on each core criteria, we run simulations. For instance, we investigate what would have happened to PCR success ratings had OVE adhered to MGTs rating for effectiveness, all else equal. These are presented in Table 5 as OVE Rating Simulations 1 to 4. We also assess how PCR scores would vary under alternative rating systems. For instance, we simulate the impact an equal weighing of 1/4 for each core criteria would have on project success[6]. That is, we run a simulation to shed light on the question *what if the IDB had adopted in its PCR Guidelines equal weights for each core criteria?* Table 5 presents these results as OVE Rating Simulations 5 and 6. When looking at these simulations, it is worth keeping in mind that MGT rated 80% of PCR17 as >=Partly Successful, and 68% for PCR18. Similarly, OVE rated 67.5% of PCR17 projects as >=Partly Successful and 50.9% for PCR18. These figures are included in the first two rows of Table 5.

As would be expected, the largest impact comes from OVE's changes to MGT's Effectiveness rating (OVE Rating Simulation 2). If OVE had adhered to MGT's effectiveness rating 56% of PCR18 projects would have been rated as >=Partly Successful, as opposed to the factual 50.9%. However, this would have still fallen well short of MGTs 68.4%. For PCR17, if OVE would have followed MGTs ratings for Effectiveness, there would have only been a 2.5% difference (77.5% vis-á-vis the factual 80%). Thus, for PCR17 most of the discrepancy in project success rating between OVE and MGT can be explained by the changes to Effectiveness. However, for PCR18 even if OVE rated effectiveness exactly as MGT did, there would still be a 12.3% difference (68.4% vs. 56.1%) This analysis suggests that for PCR18, there were other important factors at play in OVEs validation, beyond the change in effectiveness to MGTs rating, It's also suggests that core criteria ratings by OVE were more compartmentalized for PCR17 as opposed to PCR18. That is, for PCR18 validation exercise, OVEs assessment of one core criteria seems to carry over and affected the ratings of other criteria, beyond what was prescribed in the PCR Guidelines.

[6] For investment loans the PCR Guidelines assign a weight of 40% to Effectiveness and 20% to each of the other core criteria (Relevance, Efficiency, Sustainability). For Policy Based Loans (PBL), given that Efficiency is not rated, Effectiveness is 60%, with Relevance and Sustainability 20% each.

Simulations 5 and 6 present the results of simulating alternative rating schemes, including eliminating binary rules[7], weighing the four core criteria equally, or a combination of both. Of particular note is simulation 5, which attempts to mimic the rating scheme utilized by the Asian Development Bank[8]. When assessing this simulation, it is useful to keep in mind that the AsDB reports 71% of projects as successful in their most recent Development Effectiveness Review[9].

- In Simulation 5 we apply equal weights to each core criteria (1/4 for investment loans; 1/3 to PBLs[10]) and do not apply binary rules. The success of IDB projects jumps considerably to 82.5% and 68.4% for projects with CO in 2017 and 2018, respectively.
- Simulation 6 shows the results from applying equal weights, but still applying IDBs binary rules. Here we notice improvements, but not nearly as dramatic as when we drop the binary rules. The success of our projects improves to 75% and 56.1% for CO 2017 and CO2018, respectively.

[7] At the IDB, as per the PCR Guidelines, the PCR overall rating is subject to the following three rules: (i) With unsatisfactory rating for relevance, effectiveness, sustainability: highest achievable rating is partly unsuccessful; (ii) For a highly successful rating, no core-criteria rating can be less than satisfactory; and (iii). With a partly unsatisfactory rating for relevance, effectiveness or sustainability, highest achievable overall rating is partly successful

[8] see https://www.adb.org/sites/default/files/institutional-document/32516/guidelines-evaluation-public-sector.pdf

[9] See: https://www.adb.org/sites/default/files/institutional-document/602911/defr-2019.pdf

[10] It is not clear that this is the weight AsDB applies to its PBLs. AsDB guidelines refer to a qualitive assessment of efficiency that appears to be rated on equal footing as a CBA or CEA for investment loans. Given that we do not assess efficiency for PBLs, the results of this simulation should be considered with this caveat in mind.

Table 5: Results from simulation – Which PCR section is driving the difference between management and OVE scores

Variable	(1) Overall Mean	(2) Mean PCR17	(3) Mean PCR18	(4) Diff. PCR18 vs. PCR17
Mngmt. Rating >=Partly Successful	0.732	0.800	0.684	-0.116
	(0.445)	(0.405)	(0.469)	(0.092)
OVE Rating >=Partly Successful	0.577	0.675	0.509	-0.166^
	(0.497)	(0.474)	(0.504)	(0.102)
OVE Rating Simulation 1 (Relevance) >=Partly Successful	0.598	0.700	0.526	-0.174*
	(0.493)	(0.464)	(0.504)	(0.101)
OVE Rating Simulation 2 (Effectiveness) >=Partly Successful	0.649	0.775	0.561	-0.214**
	(0.480)	(0.423)	(0.501)	(0.097)
OVE Rating Simulation 3 (Efficiency) >=Partly Successful	0.588	0.700	0.509	-0.191*
	(0.495)	(0.464)	(0.504)	(0.101)
OVE Rating Simulation 4 (Sustainability) >=Partly Successful	0.577	0.700	0.491	-0.209**
	(0.497)	(0.464)	(0.504)	(0.101)
OVE Rating Simul. 5 (Equal Weights without binary rules) >=Partly Successful	0.742	0.825	0.684	-0.141^
	(0.440)	(0.385)	(0.469)	(0.090)
OVE Rating Simulation 6 (Equal Weights with binary rules) >=Partly Successful	0.639	0.750	0.561	-0.189*
	(0.483)	(0.439)	(0.501)	(0.098)
Observations	97	40	57	97

Comparing OVEs validation for PCR17 and PCR18 (Transition Matrix Analysis)

In this section we use transition matrix analysis to shed additional light on OVE's validation process from PCR17 to PCR18. Figure 6 presents these matrices. The first thing to notice is that if there had been complete agreement between MGT and OVE's rating, then only cells running through the diagonal would be filled and with the darkest shade. Comparing PCR17 to PCR18, two things jump up. First, there appears to be less agreement between MGT's rating and OVE's for PCR18. Second OVE's ratings for PCR18 appear to have greater mobility in the sense that it downgrades MGTs rating by than one grade. For instance, for PCR17 19 projects were rated as Partly Successful by MGT. Of these, OVE was in agreement for 63%. For 26% (5 PCRs), OVE downgraded the rating to Partly Unsuccessful. For two projects, it actually upgraded MGT's rating to Successful and Highly Successful. For PCR18, MGT rated 19 projects as Partly Successful. Of these OVE downgraded 43%.

The mobility index proposed in Shorrock (1978) corroborates these observations.

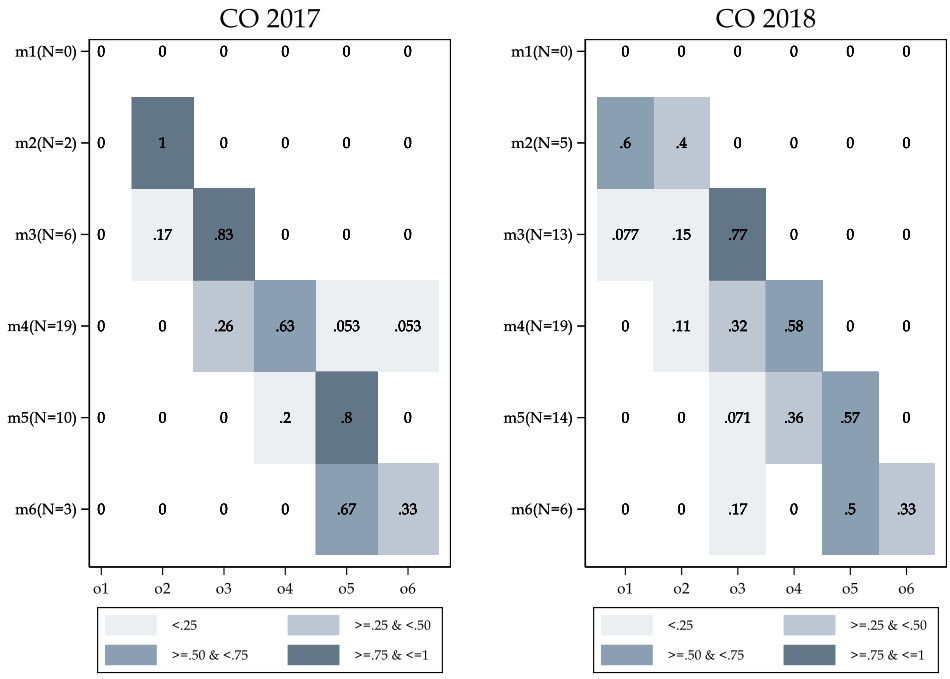
Shorrock Mobility Index, is calculated as (N-Trace(M)) /N-1.

Shorrock mobility index PCR17 – (6 – 3.59) / 5 = 0.482

Shorrock mobility index PCR18 – (6-2.65) / 5=0.67

When the index is 1, there is perfect mobility. When the index is zero, there is no mobility. Based on the index, we infer that mobility across categories increased from PCR17 to PCR18.

Figure 6: Transition Matrix - Management versus OVE rating by CO year



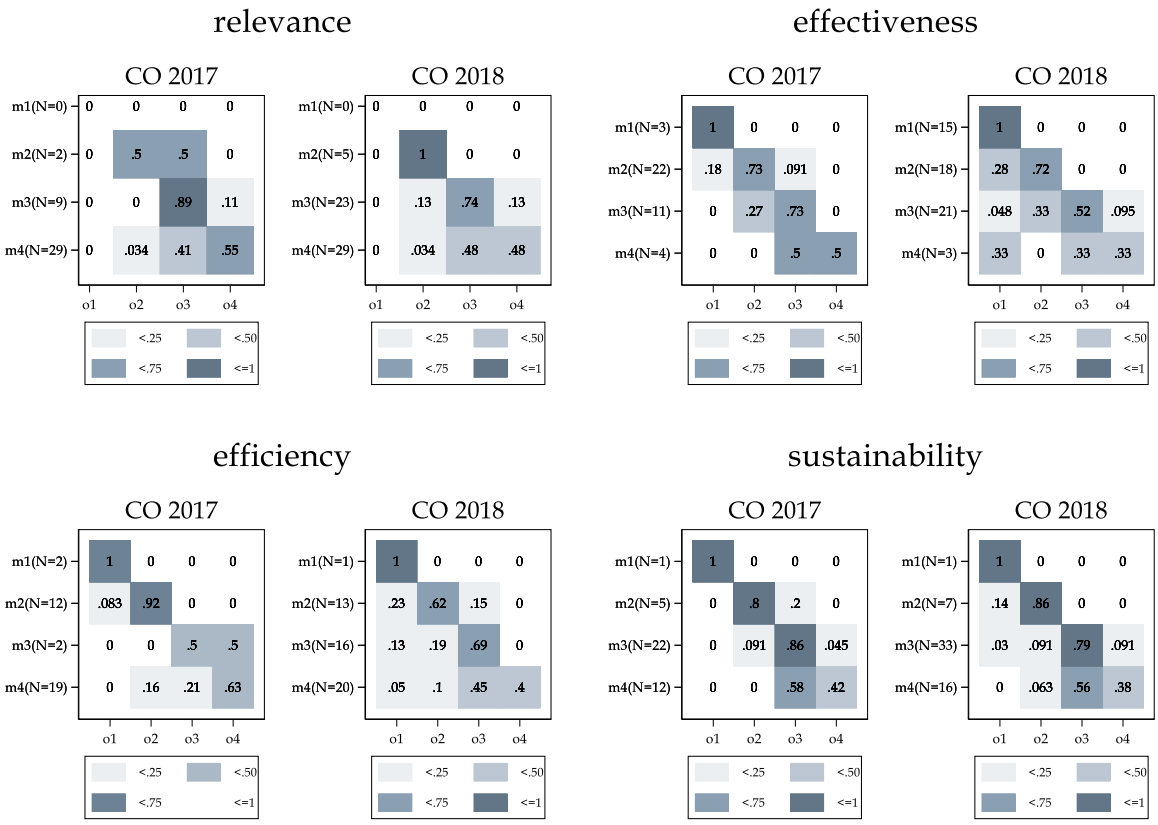
Note: m stands for management (y axis) and o for OVE (x axis). Grades goes from 1 (highly unsuccessful) to 6 (highly successful). The interpretation is by row. Each box represents a greater share of projects with the same management classification. The darker the color the greater the share.

In Figure 7 we present transition matrices by core criteria. The biggest mobility appears to have happened in Effectiveness from PCR17 to PCR18. That is, OVEs validation of Effectiveness is not just less similar to MGTs for PCR18, but it's downgrades are more pronounced. This is corroborated by the estimation of Shorrock Mobility Index -- (N-Trace(M))/(N-1) – presented in Table 6, for all core criteria. As can be seen, change in mobility between PCR17 and PCR18 is greatest for Effectiveness. The most consistent validation across PCR cycles appears to be for Sustainability. OVE agreed more with Management's rating of Relevance for PCR18 and less for Efficiency.

Table 6: Shorrock Mobility Index by PCR criteria

	All	Relevance	Effectiveness	Efficiency	Sustainability
2017	0.48	0.69	0.35	0.32	0.31
2018	0.67	0.59	0.48	0.43	0.32

Figure 7: Transition Matrix - Management versus OVE rating by CO year and PCR core criteria



Given the importance the Effectiveness rating has on overall project success and noting that the greater discrepancy between OVE and MGTs rating seems to be for this criterion, in the next section, we delve into factors associated with the effectiveness rating. Specifically, we review the 152 specific objectives of PCR18 validated by OVE, systematically identifying common issues and coding these for our analysis.

Effectiveness Analysis – by specific objectives and result indicators

IDB projects typically include a general development objective, establishing what the project aims to contribute to higher level development goals, and more narrow and attributable specific objectives. The latter describe concrete results that the project intends to achieve at completion.

The PCR evaluation methodology is objective based. The effectiveness rating of a project is based on the rate of achievement of the specific objectives, as determined by concrete results/metrics linked to each specific objective. The assessment of effectiveness of a project comprises the comparison of the intended results with the actual results and providing a convincing argument that these changes are attributable to the project's interventions.

The central task, as far as the effectiveness assessment of the PCR goes, is the establishment of the specific objectives to be evaluated and their associated results metrics that will support the demonstration of achievement. Most PCR18 projects were approved between 2009 and 2013 when, typically, projects were designed around components, not specific objectives. In practice, for this cohort of projects, establishing the specific objectives and their associated metrics is not always trivial and requires making an analytical case for their selection as well as a reasoned selection of their associated results metrics. Three main factors affecting the effectiveness rating emanate from this arrangement: design issues; execution issues; and methodological discrepancies between Management and OVE.

Take design issues first. Projects designed before August 2020, when the revised DEM came into effect[11], were not required to include at least one valid results metrics for each specific objective that the project aimed to achieve. As such, one of the first issues that can crop up is the lack of appropriate results indicators for each specific objective. Yet even when there is a valid results metrics to demonstrate achievement of the specific objective, the establishment of end of project targets might have been unrealistic or too ambitious, and thus their achievement fell short. Or, the underpinnings of the results chain were weak and hence the supply of outputs did not lead to the expected results. In some cases, SMART indicators might be absent, or they are more akin to outputs than outcomes.

During execution, outputs might have been cancelled, negatively affecting the results chain and the delivery of results. Or monitoring and evaluation systems failed to report on result indicators.

[11] The revised DEM includes two key requirements: Project proposals must include (i) at least one valid outcome indicator to assess achievement of each specific objective; and (ii) an explicit monitoring and evaluation proposal for assessing each SO as part of M&E plan proposal. In addition, a new Results Matrix format, structured around specific objectives and their results indicators, also became a requirement.

The third factor affecting the effectiveness rating relates to methodological discrepancies between Management and OVE. To start, OVE might not be in agreement with the declaration of specific objectives by Management, adding or delating as per their appreciation[12]. The same goes for the results indicators proposed by Management, including rejecting additional results indicators proposed by Management for cases where a Specific Objective lacked a valid indicator in the project's original Results Matrix. Finally, OVE might disagree with Management's arguments in support of attributing the observed changes in the results indicators to the project's interventions.

All these factors come into play in the rating of the project's effectiveness and can have significant implications for the final rating. For instance, as per the PCR Guidelines, even when a project achieves more than half of its specific objectives, but one is rated unsatisfactory the project as a whole is given a rating of Partly Unsatisfactory. That means, for example, that if OVE does not accept the results indicator proposed by Management for a specific objective as valid, or its attribution analysis, the project's effectiveness is rated Partly Unsatisfactory or below, even if the project largely achieved all its other objectives. Or, for example, take a project with two objectives. The first objective has 4 results indicators associate with it, and all but one were fully achieved. The second objective has three result indicators and all were achieved. So overall the project had 7 results indicators, all but one were fully achieved. Yet, given the rating scheme prescribed by the PCR Guidelines, this project will receive a rating of Partly Unsatisfactory for Effectiveness. Even if the project received an Excellent rating in all other criteria (Relevance, Efficiency and Sustainability), the project will end up with only a Partly Successful rating.

In this section we focus on the effectiveness analysis conducted on 62 PCR18 documents prepared and validated with comparable methodologies. On total, these 62 projects included 152 specific objectives and 515 associated result indicators, including 45 indicators that were added subsequently to provide evidence of the achievement of specific objectives that were missing valid metrics. This section is organized as follows. First, we present statistical tallies for specific objectives for sectors and country departments. We then focus on the 96 (63%) specific

[12] For instance, the objective of the National Tourism Program for Pernambuco State (BR-L1212), as per the loan proposal, states : *The objective of the operation is to increase revenue generated by tourism through development of the sun and sand model and diversification (thematic and geographic) of the tourism offerings of the state of Pernambuco*. From this statement of objectives, Management understood that the general objective was to increase revenue generated by tourism, and that the project had two specific objectives: (i) development of the sun and sand model; and (ii) diversification (thematic and geographic) of the tourism offerings of the state of Pernambuco. OVEs interpretation was that the project needed to assess the achievement of a third specific objective, namely the increase of revenue generated by tourism.

objectives that were rate unfavorably by OVE, examining the main reasons behind these ratings. Next, we delve into the achievement of results indicators, finding that nearly half of the 515 were fully attained[13]. Finally, given this apparent inconsistency between the achievement at the level of results indicators vis-à-vis at the level of specific objectives, we probe the effectiveness rating itself by simulating the impacts of changes to the prescribed effectiveness rating formula as per the PCR Guidelines.

Effectiveness at the level of Specific Objectives

As shown in Table 7, the average number of specific objectives for PCR18 was 2.5, with the average specific objective having 3.4 results indicators linked to it. PCR18 in the Caribbean country department (CCB) had the highest average number of specific objectives (2.9), and the lowest number of results indicators per specific objective (2.6). Institutions for Development (IFD) had the lowest average number of specific objectives, whereas the Social sector (SCL) had the largest average number of result indicators per specific objective.

Table 7 – PCR18, Specific Objectives and Results Indicators by country and sector

	PCRs	# Specific Objectives	# Results Indicators	Average Specific Objectives	Average Results Indicators per Specific Objective
CAN	9	21	80	2.3	3.8
CCB	8	23	60	2.9	2.6
CID	22	53	181	2.4	3.4
CSC	23	55	194	2.4	3.5
CSD	11	31	115	2.8	3.7
IFD	18	36	102	2.0	2.8
INE	16	44	135	2.8	3.1
INT	2	5	19	2.5	3.8
SCL	15	36	144	2.4	4.0
All PCRs	62	152	515	2.5	3.4

OVE validated 152 specific objectives, rating 56 (37%) favorably (=>satisfactory). As shown in Figure 8a, projects from the Caribbean had the lowest proportion of specific objectives rated favorably (30.4%), and projects in the Andean countries had the highest proportion of specific objectives rated favorably (42.8%). In terms of sector differences, Figure 8b shows the

[13] The achievement ratio used in this analysis corresponds to the one computed by OVE for its validation. It can differ from the achievement ratio reported by Management in the PCR documents and checklists.

proportion of specific objectives rated favorably by OVE ranges from 22.2% for SCL to 48.4 for CSD.[14]

Figure 8a. Number of Specific Objectives by OVE rating – by VPC Department

Figure 8b. Number of Specific Objectives by OVE rating – by VPS Department

[14] INT is not included in the Figure and additional analysis. The sector had 2 PCRs with 5 SOs (2 of them were rated favorably by OVE).

In what follows, we focus on the 96 (63%) specific objectives that were rate unfavorably by OVE (=<partly unsatisfactory). After a systematic review and coding exercise, the main reasons for not achieving specific objectives were categorized into 4 groups:

1) issues related to execution, which explain 53.1% of unfavorable rating of specific objectives.
2) issues related to project design, which explain 25% of unfavorable ratings.
3) methodological discrepancies between OVE and Management, which explain 18.8% of unfavorable ratings; and
4) other reasons, 3.1% of cases

More than half of specific objectives were rated unfavorably due to execution issues related to the cancellation of outputs (affecting the underpinning causal chain) and failing to report on the original indicators. Issues in the design phase that affected achievement of specific objectives included: i) the setting of unrealistic targets, ii) results matrix with weak causal chain or indicators that did not meet SMART criteria, and iii) specific objective had no associated metric.

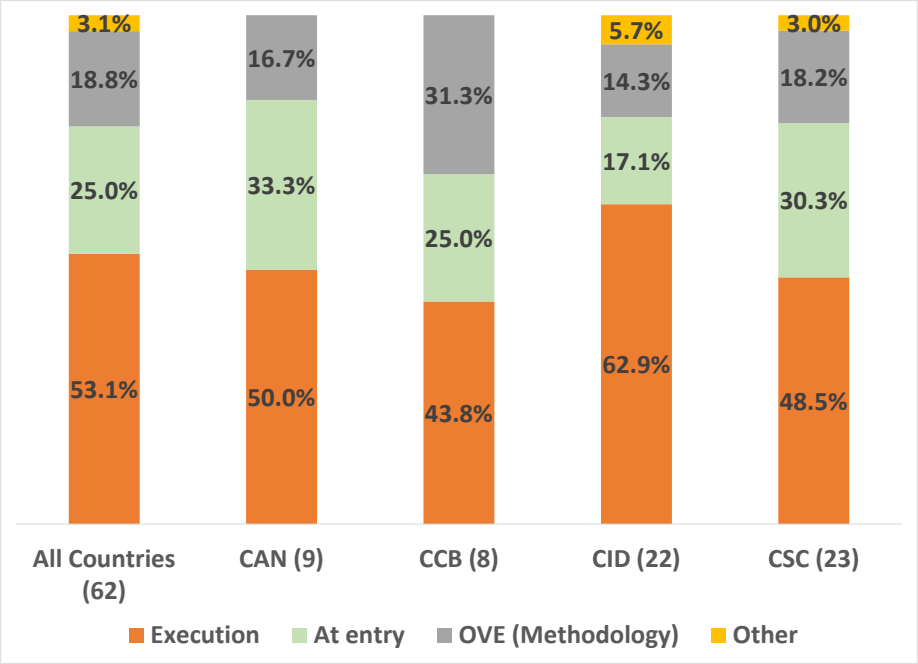
Finally, what can be categorized as methodological discrepancies between OVE and Management accounted for close to one fifth of specific objectives being rated unfavorably. The issues ranged from OVE not accepting additional indicators or the attribution analysis proposed by MGT, to OVE including or excluding indicators and adding an additional specific objective for the validation. Table 8 presents the detailed breakdown for all PCR18 and by country department and sector.

Table 8. Main Reasons for not achieving SO – detailed

	All	Country Departments				Sectors				
	PCRs	CAN	CCB	CID	CSC	CSD	IFD	INE	INT	SCL
Design Issues	**25.0%**	**33.3%**	**25.0%**	**17.1%**	**30.3%**	**18.8%**	**37.5%**	**28.0%**	**33.3%**	**14.3%**
Targets too ambitious	14.6%	16.7%	18.8%	17.1%	9.1%		25.0%	20.0%	33.3%	7.1%
Weak Results Matrix	7.3%	8.3%	6.3%		15.2%	6.3%	12.5%	4.0%		7.1%
Specific objective had no metric	3.1%	8.3%			6.1%	12.5%		4.0%		
Execution Issues	**53.1%**	**50.0%**	**43.8%**	**62.9%**	**48.5%**	**68.8%**	**37.5%**	**48.0%**	**33.3%**	**64.3%**
Cancelled outputs	24.0%	25.0%	25.0%	17.1%	30.3%	18.8%	29.2%	16.0%	33.3%	28.6%
Did not report on original indicators	24.0%	16.7%	12.5%	37.1%	18.2%	37.5%	8.3%	28.0%		28.6%
Other: Execution	5.2%	8.3%	6.3%	8.6%		12.5%		4.0%		7.1%
OVE - Methodological Discrepancies	**18.8%**	**16.7%**	**31.3%**	**14.3%**	**18.2%**	**12.5%**	**25.0%**	**16.0%**	**0.0%**	**17.9%**
Changed Results Matrix	5.2%		6.3%	2.9%	9.1%	12.5%	8.3%	4.0%		3.6%
Not accepting additional indicators	4.2%	8.3%	6.3%	2.9%	3.0%		4.2%			10.7%
Other: OVE	4.2%		12.5%	5.7%				8.0%		3.6%
Not accepting attribution analysis	3.1%		6.3%	2.9%	6.1%		8.3%	4.0%		
Adding specific objectives	2.1%	8.3%					4.2%			
Other Issues	**3.1%**			**5.7%**	**3.0%**			**8.0%**	**33.3%**	**3.6%**

Figures 9a and 9b show the distribution of these main issues leading to unfavorable ratings by country departments and sectors, respectively. The number in the parenthesis denotes the number of PCR18 in each country department or sector.

Figure 9a. Main Reasons for not achieving SO – by country department



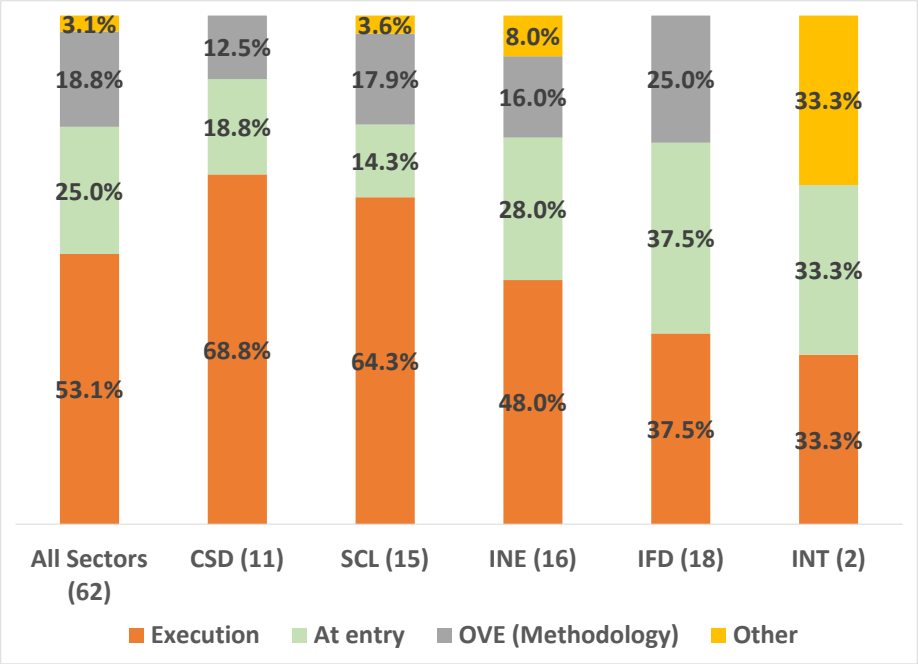
As can be gleamed in the figure above, the main reasons for not achieving specific objectives, as validated by OVE, differed across country departments, though issues related to execution dominate. For instance, for CAN and CID failures in execution account for more than half of the specific objectives that were rated unfavorably. CAN and CSC have the largest share of issues related to the design, at over 30%. Finally, CCB has the highest proportion of specific objectives rated unfavorably due to methodological discrepancies with OVE (31.5%).

With the exception of CID, the top reason for the specific objectives been rated unfavorably was cancelled outputs (25% for CAN and CCB, and 30.3% for CSC). In the case of CID and SCL the main pitfall was a failure to report on the original indicators (37.1%).

Setting unrealistic targets for results indicators is the main reason related to design issues for all, except CSC, country departments (16.7% for CAN, 18.8% for CCB, and 17.1% for CID). For CSC projects the main design issues were related to results matrix with weak causal chain or indicators that did not meet SMART criteria (15.2%).

As for methodological discrepancies between OVE and Management, when analyzing by country department, CCB stands out with 37.5% of specific objectives being rated unfavorably. Three reason shared 6.3% of the cases: not accepting additional indicators or attribution analysis, and OVE including or excluding indicators in the results matrix and in 12.5% of the cases it was related to other methodological discrepancies.

Figure 9b. Main Reasons for not achieving SO – by sector



We turn now to analyzing the main issues for failing to achieve specific objectives, as validated by OVE, by sector.[15] As per Figure 9b, what we have categorized as execution issues played by far the largest role in the failure to achieve specific objectives for CSD and SCL, accounting for 68.8% and 64.3%, respectively, of specific objectives that were not achieved. For INE and IFD design issues at entry played a proportionally larger role, with 28% and 37.5%, respectively, of specific objectives failing to be achieved due to design issues. Finally. IFD has an equal proportion of specific objectives rated unfavorably due to design and execution issues and the highest proportion due to methodological discrepancies with OVE.

[15] INT is not included in additional analysis. The sector had 2 PCRs with 5 SOs (3 of them were rated partly unsatisfactory of less by OVE). In 1 case, targets were too ambitious, another case was due to cancellation of outputs, and the third one was because of other reasons.

From Table 8, we can delve deeper into the main reasons for not achieving specific objectives related to execution failings. In the case of CSD and INE, 37.5% and 28%, respectively, of the specific objectives were not rated favorably due to failure to report on the original indicators. In the case of IFD and SCL this was related to cancelled outputs, resulting in a break of the causal chain that underpins the achievement of outcomes and specific objectives (29.2% and 28.6% respectively).

In terms of design flaws, overly ambitious or unrealistic targets caused 25% of IFD and 20% of INE's specific objectives to be rated unfavorably by OVE. Whereas for CSD not having valid metrics explains 12.5% of specific objectives rated unfavorably by OVE. Finally, in the case of SCL projects targets too ambitious and weak results matrix were equally behind unfavorable ratings, with a 7.1% share each.

As for methodological discrepancies, these played the largest role among sectors for IFD PCR18, explaining a quarter of specific objectives rated unfavorably, primarily on account of not accepting the attribution analysis (8.3%), and OVE's inclusion or exclusion of indicators in the results matrix (8.3%). The addition of specific objectives by OVE for the validation explains 4.2% of specific objectives rated unfavorably. The inclusion or exclusion of indicators was the sole source of methodological discrepancies for CSD, resulting in 12.5% of specific objectives rated unfavorably. Finally, for SCL OVE not accepting additional indicators proposed by Management resulted in 10.7% of specific objectives deemed unfavorably.

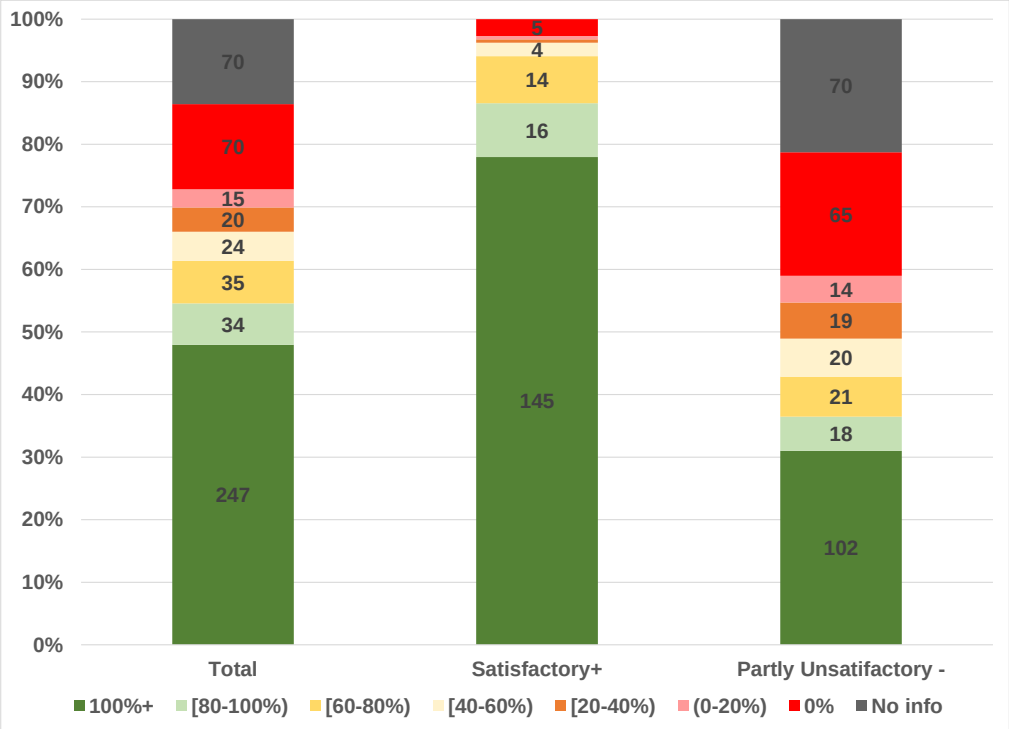
Effectiveness at the level of Result Indicators

Given that the average specific objective has 3.4 results indicators associated with it, it is feasible for a specific objective to be rated unfavorably even though some of its associated indicators met their goal. In this section we delve deeper into the effectiveness analysis turning our focus to the 515 result indicators included in the 62 PCR18 documents prepared and validated with comparable methodologies.

We begin by classifying each result indicator, in accordance with its level of achievement, from 0% to 100%. Additionally, we identified those indicators that were not measured at completion (No information). Moreover, we classify each result indicator by whether it belonged to a specific objective validated favorably (=>satisfactory) by OVE or unfavorably (=<partly unsatisfactory). The results from this exercise by Sector and Country Department are presented in Table 8 below and in percentage terms in Table A5 in the Appendix

As discussed, the 152 specific objectives validated by OVE for PCR18 were assessed using 515 result indicators. As shown in Figure 10, 48% (247) of all the result indicators were fully met, 13.6% (70) had 0% achievement ratio, and another 13.6% (70) were not measured at completion. Out of the 515 results indicators, 36.1% (186) were associated with specific objectives that were rated favorably by OVE. Notably, all the result indicators associated with specific objectives rated favorably were measured at completion. Of these 78% (145) were fully met and 2.7% (5) had 0% achievement ratio. In contrast, specific objectives rated unfavorably by OVE (63.9% of all) had 329 associated result indicators, with only 31% (102) meeting their target, 19.8% (65) had 0% achievement ratio and 21.3% (70) were not measured at completion.

Figure 10. Achievement Ratio of the Result Indicators



We turn now to heterogeneity across country departments and sectors presented in Table 8. Take country departments first. PCR18 in CSC had the highest share of fully achieved results indicators (52.6%, 102 indicators), and the lowest share of indicators that were not measured at completion (6.2%, 12 indicators). CSC also had the highest share of results indicators measuring specific objectives rated favorably (47.9%, 93 indicators). Whilst CCB had the lowest share (20%, 12 indicators). CID, followed by CAN had the highest number of indicators that were not measured due to monitoring failings, 28 (15.5%) and 11 (13.8%), respectively

In terms of sectors[16] PCR18 in CSD had the highest share of fully achieved results indicators (59.1%, 68 indicators), and the lowest share of indicators that were not measured at completion (4.3%, 5 indicators). CSD had the highest share of results indicators measuring specific objectives rated favorably (52.2%, 60 indicators). In contrast, more than 25% of the result indicators included in SCL's PCR18 were not measured at completion (27 indicators). Crucially, specific objectives that were rated unfavorably had 31% (37) indicators that were not measured. IFD had close to 20% of indicators with 0% achievement. Undoubtably, for IFD having close to a third of its result indicators with 0% achievement or not measured played an important role in that close to two thirds of its specific objectives for PCR18 were rated unfavorably.

[16] INT is not included in additional analysis. The sector had 2 PCRs with 5 SOs and 19 result indicators.

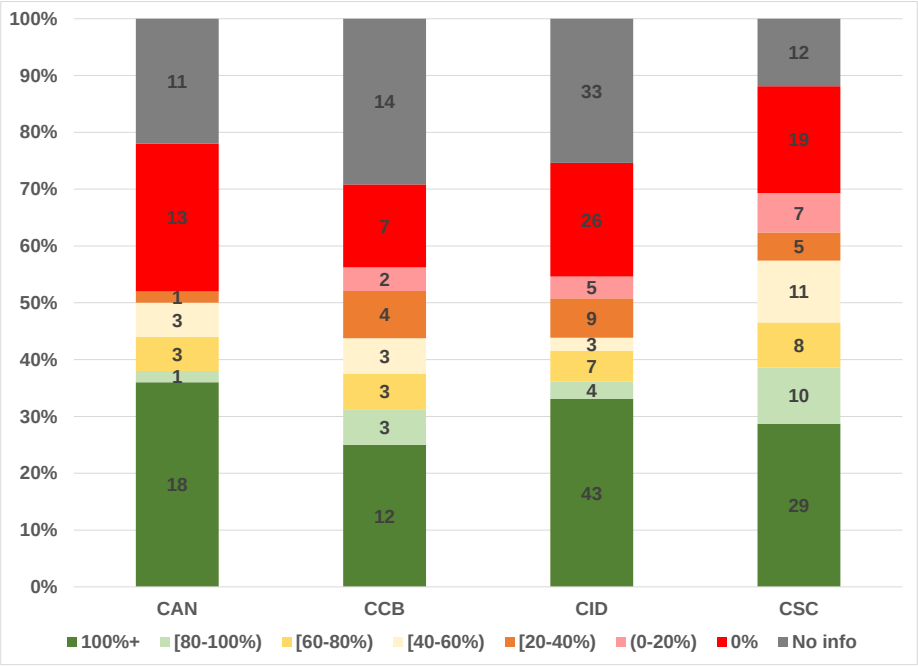
Table 9. Achievement Ratio of the Result Indicators – detailed

	All PCRs	Country Departments				Sectors				
		CAN	CCB	CID	CSC	CSD	IFD	INE	INT	SCL
PCRs #	**62**	**9**	**8**	**22**	**23**	**11**	**18**	**16**	**2**	**15**
Specific Objectives #	**152**	**21**	**23**	**53**	**55**	**31**	**36**	**44**	**5**	**36**
All result indicators	**515**	**80**	**60**	**181**	**194**	**115**	**102**	**135**	**19**	**144**
100%	247	38	19	88	102	68	43	65	12	59
[80%-100%)	34	5	6	6	17	9	10	9	1	5
[60%-80%)	35	6	4	9	16	7	7	13	1	7
[40%-60%)	24	5	4	3	12	3	4	9	1	7
[20%-40%)	20	1	4	9	6	5	4	3	1	7
(0%-20%)	15	-	2	5	8	4	2	2	1	6
0%	70	14	7	28	21	14	20	18	2	16
No information - Cancelled Outputs	19	-	12	5	2	2	-	-	-	17
No information - Monitoring	51	11	2	28	10	3	12	16	-	20
Result indicators (Satisfactory+ Specific Objectives)	**186**	**30**	**12**	**51**	**93**	**60**	**37**	**57**	**5**	**27**
100%	145	20	7	45	73	50	28	38	5	24
[80%-100%)	16	4	3	2	7	3	4	8	-	1
[60%-80%)	14	3	1	2	8	3	1	9	-	1
[40%-60%)	4	2	1	-	1	1	1	2	-	-
[20%-40%)	1	-	-	-	1	1	-	-	-	-
(0%-20%)	1	-	-	-	1	1	-	-	-	-
0%	5	1	-	2	2	1	3	-	-	1
No information	-	-	-	-	-	-	-	-	-	-
Result indicators (Partly Unsatisfactory- Specific Objectives)	**329**	**50**	**48**	**130**	**101**	**55**	**65**	**78**	**14**	**117**
100%	102	18	12	43	29	18	15	27	7	35
[80%-100%)	18	1	3	4	10	6	6	1	1	4
[60%-80%)	21	3	3	7	8	4	6	4	1	6
[40%-60%)	20	3	3	3	11	2	3	7	1	7
[20%-40%)	19	1	4	9	5	4	4	3	1	7
(0%-20%)	14	-	2	5	7	3	2	2	1	6
0%	65	13	7	26	19	13	17	18	2	15
No information - Cancelled Outputs	19	-	12	5	2	2	-	-	-	17
No information - Monitoring	51	11	2	28	10	3	12	16	-	20

In Figures 11a and 11b we show the distribution of the achievement ratios of the results indicators that measured specific objectives rated unfavorably by country departments and sectors, respectively.

Figure 11a. Achievement Ratio of the Result Indicators, by Country Department

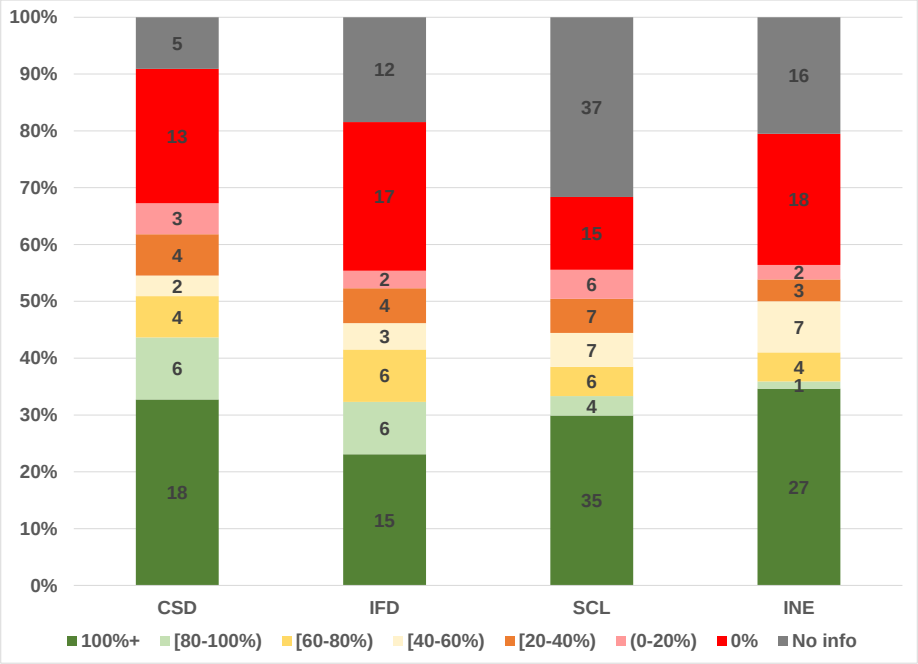
Specific Objectives Rated Partly Unsatisfactory or Unsatisfactory



The numbers included in Figure 11a represent 63.9% of all the result indicators used to measure the 152 specific objectives included in the 62 PCR18. By country department, it represents 62.5% (CAN), 80% (CCB), 71.8% (CID), and 52.1% (CSC), of all results indicators, respectively. As the figure demonstrates, it is feasible for a specific objective to be rated unfavorably even though some of its associated indicators met their goal. As the figure shows, a factor that plays into the unfavorable rating is the number of indicators for which monitoring systems did not measure achievement at completion. CCB was the country department with the highest share (29.2%, 14 indicators), followed by CID (25.4%, 33 indicators).

Figure 11b. Achievement Ratio of the Result Indicators, by sector

Specific Objectives Rated Partly Unsatisfactory or Unsatisfactory



The numbers included in Figure 11b represent 63.9% of all the results indicators used to measure the 152 specific objectives included in the 62 PCR18. By sector, it represents 47.8% (CSD), 63.7% (IFD), 57.8% INE, and 81.3% (SCL), of all results indicators, respectively.

As per figure 11b, not measuring the result indicators at completion, played a large role in the failure to achieve the specific objectives for SCL (31.6%, 37 indicators). For the rest of the sectors, indicators that achieved 0% impacted the unfavorable rating of specific objectives: 26.2% for IFD, 23.6% for CSD, and 23.1% for INE respectively.

Simulation of changes to the effectiveness rating formula

Given the apparent inconsistency between the overall effectiveness rating for PCR18, where less than 30% were rated favorably, and the finding that close to 50% of result indicators fully achieved their targets, in this section, through simulations, we assess how changes to certain parameters of the effectiveness rating formula affect the overall effectiveness rating.

For this exercise we focus on the 21 projects that received a partly unsatisfactory rating for effectiveness. The Effectiveness rating of the PCR Guidelines stablishes that a project is rated

partly unsatisfactory (PU)[17] on two accounts: (1) when more than 50% of the project's specific objectives were largely or fully achieved (and results were attributable to the project), but at least one specific objective was rated Unsatisfactory; or (2) when 50% or more of the project's specific objectives were partly achieved (or above), and the specific objectives rated Unsatisfactory do not exceed the specific objectives rated Satisfactory or Excellent.

In turn, the rating for each specific objective is dependent on the average achievement rate of its attributable associated results indicators. Thus, a specific objective whose associated result indicators' targets were fully achieved is rated Excellent (taking a value of 4); one that achieved at least 80% of the targets of the result indicators is rated Satisfactory (taking a value of 3); one that achieves between 51% and 79% of its result indicators is rated Partly Unsatisfactory (assigned a value of 2); and a specific objective that achieves 50% or less of its result indicators is rated Unsatisfactory (assigned a value of 1).

The first thing we do is attempt to replicate the rating for the 21 projects that OVE gave a rating of PU (2). For four projects that OVE rated PU, the application of the formula prescribed by the PCR Guidelines yielded a rating of Unsatisfactory (1). After reviewing OVE's validation document for these projects, we are unable to determine why this discrepancy. Regardless, in the simulations presented in what follows, none of these projects were affected by the simulated tinkering with the effectiveness formula.

The two changes to the formula we simulate are the following. First, the actual formula for a specific objective to be rated Satisfactory (3) requires that on average 80% of targets were met. This implies that the rating for a specific objective that has 4 results indicators, fully achieves 3 and 0% for the fourth, will be rated PU. In the simulations S1 and S2, presented in the in Table 10, we ask the question how many of the 21 projects that are rated PU would switch to be rated Satisfactory if the requirement for achievement was lowered from 80% to 75% or to 66%, respectively. If the requirement for a Satisfactory rating was changed to 75%, all else equal, 2 projects rated PU by OVE would be upgraded to Satisfactory. If the change was to 66%, all else equal, the effectiveness rating would change for 4 projects from PU to Satisfactory. This would imply that, under this change, 35% of PCR18 would have been rated favorably for effectiveness.

[17] In addition to PU, assigned a value of 2, Effectiveness can take on the following ratings: Excellent, assigned a value of 4, when the project's specific objectives were fully achieved; Satisfactory, assigned a value of 3, when more than 50% of the project's specific objectives were largely or fully achieved and no specific objectives are rated Unsatisfactory; and Unsatisfactory, assigned a value of 1, when the project's specific objectives were not achieved, or if the project's specific objectives have only a combination of Partly Satisfactory and Unsatisfactory ratings.

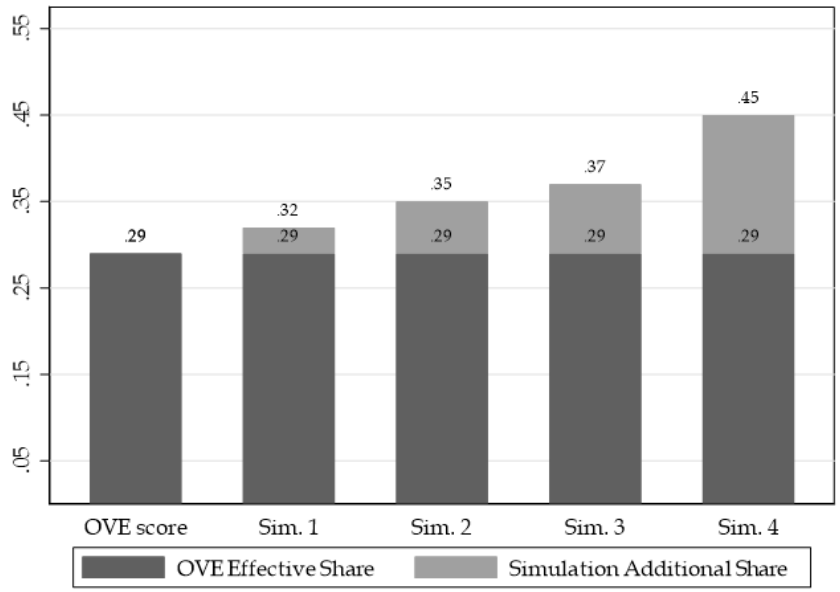
Simulations S3 and S4 present similar simulations to S1 and S2, except that in addition we do away with the rule that if one specific objective is rated Unsatisfactory the highest rating achievable is PU. With this additional change, when the target for result indicators is set to 75%, 23 projects would have been rated favorably for effectiveness by OVE. When the target is set to 66%, 28 projects would have been rated favorably for effectiveness by OVE, moving the overall favorable rating for effectiveness for PCR18 to 45%.

Table 10. Simulations to changes to the Effectiveness rating formula.

	OVE scores	S1 (C1=0.75; C2=0.5; 1 Unsat. Obj. maps to not effective)	S2 (C1=0.66; C2=0.5; 1 Unsat. Obj. maps to not effective)	S3 (C1=0.75; C2=0.5)	S4 (C1=0.66; C2=0.5)
N	62	62	62	62	62
N Evaluation Score (>=3)	18	20	22	23	28
% Evaluation Score (>=3)	0.29	0.32	0.35	0.37	0.45

Figure 12 presents the result of these simulation graphically.

Figure 12. Simulations to changes to the Effectiveness rating formula



Note: Simulation (Sim.) 1 assumes cutoff1=0.75 with dichotomic rule for the presence of 1 unsatisfactory specific objective; Sim. 2 assumes cutoff1=0.66 with dichotomic rule for the presence of 1 unsatisfactory specific objective; Sim. 3 assumes cutoff1=0.75 without the dichotomic rule; and Sim. 4 assumes cutoff1=0.66 without the dichotomic rule.

Determinants of Project Effectiveness and Overall Success

In this section we will attempt to assess the role the DEF, and other factors, play in determining a project's effectiveness and overall success as captured by PCR ratings. We will attempt to explain for both MGT and OVE scores what determines (or is associated with) whether a project is rated successful or not (and effective or not). The explanatory variables include the quality of the project at entry (DEM), its performance in execution (PMR), and other characteristics such as sector, country, modality (multiple works, specific investment, PBL), etc. For this, we use linear probability models (LPM). LPM impose less restrictive distributional assumptions but can generate probabilities outside the unit interval. For common values of the covariates, the linear probability often approximates the response probability (Wooldridge 2002). An advantage of using LPM lies in that estimates are directly interpretable as marginal effects, whereas probit and logit coefficients must be transformed before they can be interpreted as such. An anticipated limitation to the robustness of this analysis lies with the relatively few observations and thus degrees of freedom.

Characteristics of Successful and Effective Projects

We start out by analyzing project characteristics in design and execution associated with more successful projects and more effective projects as rated by OVE and MGT.

Table 11 presents quality at entry characteristics, as evaluated by the DEM, according to whether a PCR was rated Partly Successful or higher or less than Partly Successful by OVE. Overall, successful projects exhibited higher DEM scores in all dimensions, and significantly so with regards to quality of the ex-ante economic analysis, the result matrix, and the outcome indicators. In order to assess whether the complexity of a project's design is associated with project success, we created 3 variables from the PMR data base to attempt to capture project complexity. These include: (i) the number of outcomes over the number of components; (ii) the number of outputs over the number of components; and (iii) the number of components. We find evidence that more successful projects are in general less complex, as proxied by these measures. Particularly, with regards to number of components.

Table 12 presents execution performance indicators, constructed from the PMR, according to whether a PCR was rated Partly Successful or higher or less than Partly Successful by OVE. Overall, successful projects outperformed their unsuccessful peers in nearly every execution aspect. Successful projects disbursed a higher share from the approved amount (nearly 13% higher) and were less likely to have a large percent of the original amount canceled. On average, successful projects executed in less time, closing 6 months earlier. Successful projects

were much less likely to be put in alert or classified as a problem project, both early on (first 3 years) and later in its execution life (last 3 years), and they were less likely to be in alert on their last year of execution. Finally, successful projects had on average a significant lower share of outputs from the first results matrix entered into the system that were discontinued or deactivated (8.5% less). Successful projects also, on average, had a significant lower share of outputs that were not achieved (10.6%). This result suggests that the more a project is able to adhere to its original vertical logic, the more likely it will be successful. On the whole, these findings suggest execution performance is highly influential on project success and that it might be a necessary condition, albeit not a sufficient one, to assure project success.

Table 13 presents quality at entry characteristics, as evaluated by the DEM, according to whether a PCR was rated as satisfactory in Effectiveness by OVE. The DEM score for projects rated satisfactory or above for Effectiveness is significantly higher, as is the quality of the results matrix and of its outcome indicators and the quality of its Monitoring and Evaluation Plan, driven by the robustness of the evaluation proposed at entry. Credibly demonstrating the attribution of results in the effectiveness analysis is a task that is greatly facilitated with a robust evaluation. Finally, successful projects are more likely to have fewer number of components.

Table 14 presents execution performance indicators by Effectiveness rating by OVE. Projects with superior execution performance were more likely to achieve their original specific objectives and their associated results, as captured by a satisfactory rating in Effectiveness. Effective projects disbursed a higher share from the approved amount (9.8% higher) and did not have a large portion of the original loan amount canceled. On average, effective projects executed in less time, closing 7 months earlier. Effective projects were less likely to be put in alert or classified as a problem project, both early on (first 3 years) and later in its execution life (last 3 years), and they were less likely to be in alert on their last year of execution. They were also less likely to need extensions. Finally, effective projects experienced a lower share of outputs, from the first results matrix entered into the system, that were discontinued or deactivated (11.7% less), and they had a higher share of outputs achieved in the end. This suggests that a project that was less affected in the first links of its vertical logic, through less changes in the outputs that supported the causal chain, were more likely to achieve their objectives.

Tables 15 to 18 present the equivalent of Tables 11 to 14, but for MGT ratings.

All though higher DEM scores are associated with project success as rated by MGT, only for RM quality is the difference significant. Successful projects, as assessed by MGT, presented a higher number of outcomes per component, but fewer number of components.

Successful projects, as rated by MGT, disbursed a higher share from the approved amount (14.2% higher) and were less likely to have a large percent of the original amount canceled. On average, successful projects executed in less time, closing 5.4 months earlier. Successful projects were significantly less likely to be put in alert or classified as a problem project, both early on (first 3 years) and later in its execution life (last 3 years), and they were less likely to be in alert on their last year of execution. Successful projects, as rated by MGT, had a lower ratio of outputs that were not achieved (15.4%), and this was also reflected in the cost share for these outputs. These characteristics associated with a successful project as rated by MGT are similar to those for OVE, but MGT's effects are a bit larger.

Effective projects, as per MGT rating, are related to higher DEM scores, although only Evaluation is significantly higher. Effective projects as rated by MGT have on average a lower number of components.

In terms of execution performance and its effect in the rating of Effectiveness, the patterns are similar for MGT rating as those found for OVE. However, in general, the differences are more attenuated, with the exception of a stronger effect for a project being classified in alert or problem, particularly in its last three years of execution, which is stronger for MGT rating. MGT's rating also seems to be more affected by the share of outputs that were not achieved and their cost share.

Econometric Analysis

Having identified quality at entry and execution performance characteristics associated with superior Successful and Effectiveness ratings by OVE and MGT, in this section we turn to regression analysis to shed additional light on the marginal contribution of each of these characteristics.

Table 19 presents regression results for OVE's rating of success and effectiveness and for the OVE PCR Score. Linear probability models are employed to estimate coefficients presented in columns (1), (2), (4) and (5). LPM impose less restrictive distributional assumptions but can generate probabilities outside the unit interval. For common values of the covariates, the linear probability often approximates the response probability (Wooldridge 2002). An advantage of using LPM lies in that estimates are directly interpretable as marginal effects, whereas probit

and logit coefficients must be transformed before they can be interpreted as such. OLS is used in the estimations presented in columns (3) and (6). The only difference between columns (1) to (3) and (4) to (6) is the inclusion of sector dummies in the latter, with INE being the comparator. In general, all models have good explanatory power, with slightly higher R-squares for models that include sector dummies. For these regressions, PBLs were not included, as they do not report execution performance indicators in the PMR. Thus, these are based on 83 observations. Despite the relatively small number of observations, several coefficients are highly significant and are precisely estimated (low robust standard errors). Also, results are fairly stable across different model specifications.

In terms of project characteristics at entry, the first thing of note is the importance of the quality of the results matrix at entry in explaining project success and effectiveness. It is also strongly associated with a higher PCR score. The quality of the Evaluation Plan is associated with a project's effectiveness. As discussed above, this suggests that a robust Evaluation Plan is likely to facilitate the demonstration of attribution, thus enhancing effectiveness.

In terms of execution performance, the importance of a higher disbursement share is unequivocally confirmed. Similarly, the less a project spends in alert or classified as a problem project in its first three years of execution, the more likely it will be successful. Similarly, the more a project adheres to the execution of outputs included in the first matrix registered in PMR, the more likely it will be successful and effective. The higher the cost share from outputs that are not achieved at the end of execution, the less likely the project will be rated as effective by OVE.

Table 19. Regression Results (OVE)

	(1) OVE Rating >=Partly Successful	(2) OVE effectiveness (>=satisfactory)	(3) OVE PCR Score	(4) OVE Rating >=Partly Successful	(5) OVE effectiveness (>=satisfactory)	(6) OVE PCR Score
All - Original DEM score	0.031	-0.044	0.008	0.021	-0.074	0.007
	(0.07)	(0.06)	(0.09)	(0.07)	(0.06)	(0.09)
RM quality	0.218^	0.133^	0.267***	0.224*	0.134^	0.269***
	(0.13)	(0.09)	(0.10)	(0.13)	(0.09)	(0.10)
Evaluation	-0.042	0.089**	-0.015	-0.051	0.096**	-0.034
	(0.05)	(0.04)	(0.06)	(0.05)	(0.05)	(0.06)
N. components	-0.069^	-0.080**	-0.045	-0.085*	-0.097**	-0.061
	(0.05)	(0.04)	(0.05)	(0.05)	(0.04)	(0.05)
Disbursed share from approved amount	0.764**	0.633***	1.012***	0.739*	0.624**	0.978**
	(0.38)	(0.23)	(0.38)	(0.39)	(0.24)	(0.39)
N. months overrun	-0.006	-0.003	-0.012**	-0.006	-0.002	-0.013*
	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)
Performance Index - Alert/Problem Ratio 3 first years	-0.341*	0.027	-0.351*	-0.351*	-0.006	-0.361*
	(0.17)	(0.16)	(0.19)	(0.18)	(0.15)	(0.20)
Performance Index - Alert/Problem Ratio 3 last years	-0.033	0.006	0.019	-0.011	0.033	0.041
	(0.23)	(0.16)	(0.25)	(0.23)	(0.16)	(0.26)
Share outputs 1st matrix discontinued	-0.322^	-0.329**	-0.412^	-0.338^	-0.354**	-0.436*
	(0.22)	(0.16)	(0.27)	(0.23)	(0.17)	(0.26)
Cost share from outputs last matrix not achieved	-0.137	-0.336**	-0.156	-0.139	-0.362**	-0.145
	(0.17)	(0.13)	(0.17)	(0.18)	(0.14)	(0.19)
1[Year Approval>=2012]	0.021	-0.042	-0.064	0.035	0.042	-0.076
	(0.18)	(0.13)	(0.18)	(0.19)	(0.16)	(0.20)

CSD				0.028	0.036	0.061
				(0.18)	(0.17)	(0.22)
IFD				-0.108	-0.204^	-0.054
				(0.16)	(0.13)	(0.16)
INE				0	0	0
				(.)	(.)	(.)
SCL				0.066	-0.107	0.158
				(0.18)	(0.18)	(0.21)
Constant	-0.288	-0.297	1.466**	-0.091	0.029	1.623**
	(0.66)	(0.57)	(0.70)	(0.68)	(0.56)	(0.72)
R-squared	0.316	0.259	0.323	0.331	0.298	0.337
N. of cases	83	83	83	83	83	83

Table 20 presents the equivalent results for MGTs rating of success and effectiveness. In general, there seems to be some level of coincidence with determinants of success as rated by OVE. Results matrix quality is the only DEM dimension that is a significant determinant of project success as rated by MGT.

In terms of execution performance, similar to OVE, a higher disbursement share is associated with a higher likelihood of being rated successful by MGT. Similarly, the less a project spends in alert or classified as a problem project in its first three years of execution, the more likely it will be rated successful by MGT. And the more a project adheres to the execution of outputs included in the first matrix registered in PMR, the more likely it will be rated effective by MGT. The higher the cost share from outputs that are not achieved at the end of execution, the less likely the project will be rated as successful or effective by MGT

Table 20. Regression Results (MGT)

	(1) MNGMNT Rating >=Partly Successful	(2) MNGMNT effectiveness (>=satisfactory)	(3) MNGMNT PCR Score	(4) MNGMNT Rating >=Partly Successful	(5) MNGMNT effectiveness (>=satisfactory)	(6) MNGMNT PCR Score
All - Original DEM score	0.045	-0.016	0.03	0.048	-0.046	0.03
	(0.07)	(0.07)	(0.08)	(0.07)	(0.07)	(0.08)
RM quality	0.199*	0.045	0.160^	0.172^	0.017	0.166*
	(0.12)	(0.12)	(0.10)	(0.12)	(0.12)	(0.10)
Evaluation	-0.059	0.051	-0.015	-0.056	0.053	-0.029
	(0.05)	(0.05)	(0.06)	(0.05)	(0.06)	(0.06)
N. components	-0.043	-0.067	-0.074^	-0.036	-0.085^	-0.085^
	(0.04)	(0.06)	(0.05)	(0.05)	(0.06)	(0.05)
Disbursed share from approved amount	0.532*	0.235	0.780**	0.545*	0.224	0.756**
	(0.30)	(0.33)	(0.34)	(0.32)	(0.35)	(0.35)
N. months overrun	-0.007^	-0.004	-0.007^	-0.006	-0.003	-0.007^
	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)
Performance Index - Alert/Problem Ratio 3 first years	-0.321**	-0.054	-0.424**	-0.353**	-0.13	-0.422*
	(0.15)	(0.18)	(0.21)	(0.16)	(0.17)	(0.22)
Performance Index - Alert/Problem Ratio 3 last years	-0.172	-0.234	-0.11	-0.178	-0.201	-0.096
	(0.25)	(0.19)	(0.24)	(0.25)	(0.18)	(0.25)
Share outputs 1st matrix discontinued	-0.156	-0.306^	-0.376^	-0.187	-0.378*	-0.385^
	(0.19)	(0.20)	(0.25)	(0.19)	(0.21)	(0.24)
Cost share from outputs last matrix not achieved	-0.342**	-0.420***	-0.351**	-0.343*	-0.444***	-0.343*
	(0.17)	(0.14)	(0.18)	(0.18)	(0.15)	(0.18)
1[Year Approval>=2012]	-0.034	-0.073	-0.061	-0.014	0.034	-0.076
	(0.14)	(0.19)	(0.17)	(0.16)	(0.21)	(0.19)
CSD				0.108	0.183	0.017
				(0.14)	(0.19)	(0.16)
IFD				0.084	-0.159	-0.043
				(0.13)	(0.16)	(0.15)

INE				.	.	.
				(.)	(.)	(.)
SCL				-0.047	-0.095	0.114
				(0.16)	(0.21)	(0.20)
Constant	0.071	0.41	2.165***	0.035	0.803	2.261***
	(0.53)	(0.66)	(0.66)	(0.58)	(0.67)	(0.65)
R-squared	0.355	0.194	0.366	0.369	0.246	0.375
N. of cases	83	83	83	83	83	83

Conclusions and Recommendations

The goal of development assistance is to deliver results. At the IDB, the DEF was adopted to enhance the likelihood that this goal is met. The instruments of the DEF (DEM, PMR and PCR) were designed with the hope that each would contribute to assure that projects reach their development objectives.

This analysis has been spurred by the lackluster performance of PCR18 in relation to PCR17. But even if we consider PCR18 an anomaly, the overall effectiveness of projects is well below what one would hope for. Thus, in this report we have attempted to shed light on factors affecting project effectiveness and success.

In terms of the validation process, comparing PCR17 to PCR18, two aspects are noteworthy. First, there was greater divergence between MGT's rating and OVE's for PCR18. Second OVE's ratings for PCR18 show greater mobility, in the sense that the disagreement in ratings for specific criteria between MGTs and OVE is greater than for PCR17. We also find some evidence consistent with OVE's ratings being more compartmentalized for PCR17 as opposed to PCR18. That is, there appears to be greater correlation among ratings from one criterion to the next for PCR18, resulting in a negative rating for one criterion influencing the rating of the remaining criteria, beyond what the PCR guidelines prescribe.

With regards to the analysis of OVE's assessment of the achievement of specific objectives, the main deficiencies associated with close to 64% of specific objectives being rated unfavorably can be categorized as 1) issues related to execution, which explain 53.1% of unfavorable rating of specific objectives; 2) issues related to project design, which explain 25% of unfavorable rating of specific objectives; 3) methodological discrepancies between OVE and Management, which explain 18.8% of unfavorable rating of specific objectives; and 4) other various reasons. Execution issues included cancelation of outputs and monitoring pitfalls that did not allow for reporting on the original indicators. Issues in the design phase that affected achievement of specific objectives included: i) unrealistic targets, ii) weak results matrix, and iii) specific objective with no metrics to value achievement. Finally, methodological discrepancies between OVE and Management included: (i) OVE not accepting additional indicators or the attribution analysis; (ii) OVE altering the results matrix by including or excluding indicators; and (iii) OVE adding an specific objective that was not originally included for the validation.

We also find that despite the low achievement at the level of specific objectives, 48% of result indicators for PCR18 were fully achieved. However, nearly 10% of result indicators had no

information of achievement, indicating serious monitoring failings. We do find that relatively small changes to the Effectiveness rating formula can increase the overall effectiveness of PCR18 projects, closing the gap between achievement at the level of specific objectives and result indicators.

Overall, our findings provide strong validation for the DEF and its tools. We find robust evidence for the importance of the quality at entry assessment conducted through the DEM. In particular, the quality of the results matrix at approval is a strong determinant of project success and effectiveness at closure, and the quality of the Evaluation Plan is strongly associated with the effectiveness of a project. In terms of execution performance, our analysis finds that, on average, projects that execute a lower share of their approved loan amount; are put on Alert, or are classified as a Problem, in its first three years of execution; and experience a higher share of their outputs discontinued, with respect to their first results matrix, are most likely to be ineffective in achieving their objectives and will likely be rated as unsuccessful.

At the IDB, there has been a concerted effort to adopt management tools geared towards achievement of results. This is the case with the ex-ante quality at entry focus and at project closure with the PCR. Execution performance is still largely focused on tracking the achievement of products at cost and schedule. Although a satisfactory execution of products is a likely a necessary condition to achieve development results, it is far from sufficient. In line with the design of projects and the assessment at the closure of projects, the IDB needs to move from culture focused on achieving product in time and at cost during execution to a culture and mindset of achieving results. This will require addressing the incentives structure during execution, as well as a greater conscientization at all levels. For instance, a measure that could be taken is to include explicit review of projects' results progress during portfolio reviews, mid-term evaluations, etc. At a minimum, these reviews should include: (i) an assessment of the continued soundness of the vertical logic; (ii) a review of the existing capacity to monitor and report achievement for each result indicator; and (iii) assuring that each specific objective has a valid associated result indicator. The IDB can also adopt specific measures to counteract where the risk of a project ending unsuccessful/ineffective is higher than a pre-defined threshold. For instance, if a project cancels over 30% of its loan amount, a trigger for an automatic

reformulation can be activated[18]. We have migrated to a results-based management system at entry and closure, perhaps it is time to move in that direction during execution.

This analysis has helped identify characteristics of projects that are important in their achievement of development goals. Is has also identified characteristics that are associated with an increased risk of projects failing to reach their objectives, such as the cancellation of funds, discontinuation of products, and being put in alert or classified as problem projects. These findings can be used to analyze the current portfolio of projects in execution to identify projects that, given their characteristics and stage of execution, are at a high risk of failing to achieve their objectives. IDB Management can use this flagging to take concrete action, such as reformulation, to enhance the probability that the project will succeed. An investigation in this regard can be found in Alvarez et. al. (2021).

The analysis presented in Alvarez et al. (2021) suggests that assuring the highest quality during the design stage, including the quality of the results matrix, the evaluation plan, and limiting undue complexity can ameliorate somewhat the effects of poor execution performance, but not sufficiently to overcome them. Thus, in line with the findings of this report, in the design stage aspects that should be paid close attention to include: (i) assuring that each specific objective has valid result indicators; (ii) that result indicators' targets are realistic and calibrated to local realities; and (ii) that the attribution analysis to be conducted is pre-specified for each specific objective. The revised DEM, applied since august of 2020, includes the review of these aspects.

These new DEM requirements will go a long way to limit the main discrepancies with OVE in future. However, for older projects in the portfolio, some steps could be taken to limit these. For instance, before the PCR team begins preparation in earnest, a pre-agreement with OVE could be sought on the specific objectives that will be the basis for the evaluation and their associated result indicators used to establish achievement.

Finally, this analysis has been hampered by the number of PCRs available for inclusion. It is recommended that as more PCRs are concluded and validated, this type of analysis be replicated to obtain more robust statistical estimates and assess the role of a greater number of factors and variables in project success and effectiveness. In particular, extrapolating the findings of the PCR analysis to the portfolio under execution, as conducted in Alvarez et al. (2021) will benefit significantly from additional statistical robustness.

[18] As per analysis presented in IDB-TN-02136 - PCR Analysis: Implications for the Portfolio, a cancellation of 30% reduces the likelihood of a project closing as successful to around 20%.

References

Alvarez, C., Leonardo Corral, Ana Cuesta, José Martinez, & Cesar Montiel. (2021). PCR Analysis: Implications for the Portfolio. Inter-American Development Bank IDB-TN-02136.

Asian Development Bank. (2015). 2014 development effectiveness review. Mandaluyong City: Author. Retrieved from https://www.adb.org/documents/development-effectiveness-review-2014-report

Carneiro, B., & Garbero, A. (2017). Supporting impact with evidence: A content analysis of project completion reports. The Journal of Development Studies. doi:10.1080/00220388.2017.1324148

Corral, L., & Nancy Mccarthy (2018): Organisational Efficiency or Bureaucratic Quagmire: Do Quality-At-Entry Assessments Improve Project Performance? The Journal of Development Studies, DOI: 10.1080/00220388.2018.1554210

Crespo, A., Hallberg, K., Ruiz, N., Martin, L., Alonso, D., Zambrano, P., & Dal-Toe, T. (2013). IDB-9: The development effectiveness framework and overview. Washington, DC: Inter-American Development Bank. Retrieved from https://publications.iadb.org/handle/11319/5869

International Fund for Agricultural Development. (2014). Report on IFAD's development effectiveness. Rome, Italy: IFAD. Retrieved from https://www.ifad.org/web/guest/document-detail/asset/40689374

Shorrocks, A. F. (1978). The measurement of mobility. Econometrica: Journal of the Econometric Society, 1013-1024

United Nations Development Programme. (2001). Development effectiveness: Review of evaluative evidence. New York, NY: Author. Retrieved from http://web.undp.org/evaluation/documents/der2001.pdf

Wooldridge, J. M. (2002). Econometric analysis of cross section and panel data. Cambridge, MA: MIT Press.

World Bank. (2005). Improving the world bank's development effectiveness: What does evaluation show? Washington, DC. Retrieved from https://openknowledge.worldbank.org/handle/10986/7225

Table 1. Descriptive Statistics PCRs by CO Year, Sector and Department

Variable	(1) Overall Mean	(2) Mean PCR17	(3) Mean PCR18	(4) Diff. PCR18 vs. PCR17
CSD	0.175	0.200	0.158	-0.042
	(0.382)	(0.405)	(0.368)	(0.079)
IFD	0.351	0.425	0.298	-0.127
	(0.480)	(0.501)	(0.462)	(0.099)
INE	0.237	0.225	0.246	0.021
	(0.428)	(0.423)	(0.434)	(0.089)
INT	0.031	0.025	0.035	0.010
	(0.174)	(0.158)	(0.186)	(0.036)
SCL	0.206	0.125	0.263	0.138*
	(0.407)	(0.335)	(0.444)	(0.083)
CAN	0.206	0.300	0.140	-0.160*
	(0.407)	(0.464)	(0.350)	(0.083)
CCB	0.093	0.075	0.105	0.030
	(0.292)	(0.267)	(0.310)	(0.060)
CID	0.340	0.325	0.351	0.026
	(0.476)	(0.474)	(0.481)	(0.099)
CSC	0.361	0.300	0.404	0.104
	(0.483)	(0.464)	(0.495)	(0.100)
PBL_d	0.124	0.125	0.123	-0.002
	(0.331)	(0.335)	(0.331)	(0.069)
Observations	97	40	57	97

Table 2. Descriptive Statistics PCR Ratings by CO year

Variable	(1) Overall Mean	(2) Mean PCR17	(3) Mean PCR18	(4) Diff. PCR18 vs. PCR17
Mngmt. Rating = Highly Successful	0.093	0.075	0.105	0.030
	(0.292)	(0.267)	(0.310)	(0.060)
Mngmt. Rating = Successful	0.247	0.250	0.246	-0.004
	(0.434)	(0.439)	(0.434)	(0.090)
Mngmt. Rating = Partly Successful	0.392	0.475	0.333	-0.142^
	(0.491)	(0.506)	(0.476)	(0.101)
Mngmt. Rating = Partly Unsuccessful	0.196	0.150	0.228	0.078
	(0.399)	(0.362)	(0.423)	(0.082)
Mngmt. Rating = Unsuccessful	0.072	0.050	0.088	0.038
	(0.260)	(0.221)	(0.285)	(0.054)
Mngmt. Rating = Highly Unsuccessful	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
OVE = Highly Successful	0.041	0.050	0.035	-0.015
	(0.200)	(0.221)	(0.186)	(0.041)
OVE = Successful	0.227	0.275	0.193	-0.082
	(0.421)	(0.452)	(0.398)	(0.087)
OVE = Partly Successful	0.309	0.350	0.281	-0.069
	(0.465)	(0.483)	(0.453)	(0.096)
OVE = Partly Unsuccessful	0.289	0.250	0.316	0.066
	(0.455)	(0.439)	(0.469)	(0.094)
OVE = Unsuccessful	0.093	0.075	0.105	0.030
	(0.292)	(0.267)	(0.310)	(0.060)
OVE = Highly Unsuccessful	0.041	0.000	0.070	0.070*
	(0.200)	(0.000)	(0.258)	(0.041)
Mngmt. Rating >=Partly Successful	0.732	0.800	0.684	-0.116
	(0.445)	(0.405)	(0.469)	(0.092)
OVE Rating >=Partly Successful	0.577	0.675	0.509	-0.166^
	(0.497)	(0.474)	(0.504)	(0.102)
Mngmt. relevance (>=satisfactory)	0.928	0.950	0.912	-0.038
	(0.260)	(0.221)	(0.285)	(0.054)
Mngmt. effectiveness (>=satisfactory)	0.402	0.375	0.421	0.046
	(0.493)	(0.490)	(0.498)	(0.102)
Mngmt. efficiency (>=satisfactory)	0.711	0.650	0.754	0.104
	(0.455)	(0.483)	(0.434)	(0.094)
Mngmt. sustainability (>=satisfactory)	0.856	0.850	0.860	0.010

	(0.353)	(0.362)	(0.350)	(0.073)
OVE relevance (>=satisfactory)	0.887	0.950	0.842	-0.108^
	(0.319)	(0.221)	(0.368)	(0.065)
OVE effectiveness (>=satisfactory)	0.299	0.350	0.263	-0.087
	(0.460)	(0.483)	(0.444)	(0.095)
OVE efficiency (>=satisfactory)	0.619	0.575	0.649	0.074
	(0.488)	(0.501)	(0.481)	(0.101)
OVE sustainability (>=satisfactory)	0.794	0.825	0.772	-0.053
	(0.407)	(0.385)	(0.423)	(0.084)
Mngmt. relevance score	3.526	3.675	3.421	-0.254*
	(0.631)	(0.572)	(0.653)	(0.128)
Mngmt. effectiveness score	2.289	2.400	2.211	-0.189
	(0.853)	(0.778)	(0.901)	(0.176)
Mngmt. efficiency score	3.094	3.086	3.100	0.014
	(0.946)	(1.067)	(0.863)	(0.210)
Mngmt. sustainability score	3.124	3.125	3.123	-0.002
	(0.696)	(0.723)	(0.683)	(0.144)
OVE relevance score	3.237	3.375	3.140	-0.235*
	(0.642)	(0.586)	(0.667)	(0.131)
OVE effectiveness score	2.052	2.225	1.930	-0.295^
	(0.870)	(0.800)	(0.904)	(0.178)
OVE efficiency score	2.694	2.800	2.620	-0.180
	(0.976)	(1.052)	(0.923)	(0.216)
OVE sustainability score	2.907	2.950	2.877	-0.073
	(0.693)	(0.639)	(0.734)	(0.144)
Changes to PU due to binary rules (%)	0.062	0.000	0.105	0.105**
	(0.242)	(0.000)	(0.310)	(0.049)
Changes to PS due to binary rules (%)	0.165	0.300	0.070	-0.230***
	(0.373)	(0.464)	(0.258)	(0.074)
Overall successful downgrade	0.155	0.125	0.175	0.050
	(0.363)	(0.335)	(0.384)	(0.075)
Relevance satisfactory downgrade	0.052	0.025	0.070	0.045
	(0.222)	(0.158)	(0.258)	(0.046)
Effectiveness satisfactory downgrade	0.124	0.075	0.158	0.083
	(0.331)	(0.267)	(0.368)	(0.068)
Efficiency satisfactory downgrade	0.113	0.075	0.140	0.065
	(0.319)	(0.267)	(0.350)	(0.066)
Sustainability satisfactory downgrade	0.072	0.050	0.088	0.038
	(0.260)	(0.221)	(0.285)	(0.054)

Mngmt. PCR Score	2.847	2.920	2.796	-0.124
	(0.551)	(0.534)	(0.561)	(0.113)
OVE PCR Score	2.581	2.710	2.491	-0.219*
	(0.587)	(0.502)	(0.629)	(0.120)
OVE PCR Score - Mngmt. Score	-0.266	-0.210	-0.305	-0.095^
	(0.346)	(0.323)	(0.359)	(0.071)
Observations	97	40	57	97

Table 3. Descriptive Statistics PCRs by CO Year, DEM Scores and project characteristics.

Variable	(1) Overall Mean	(2) Mean PCR17	(3) Mean PCR18	(4) Diff. PCR18 vs. PCR17
All - Original DEM score	7.783	7.656	7.872	0.216
	(1.381)	(1.478)	(1.316)	(0.289)
S3 - Program Logic	7.902	7.785	7.984	0.199
	(1.455)	(1.379)	(1.513)	(0.304)
S4 - Economic Analysis	7.571	7.474	7.637	0.163
	(3.268)	(3.464)	(3.154)	(0.685)
S5 - M&E	7.047	6.845	7.187	0.342
	(1.903)	(1.913)	(1.901)	(0.398)
RM quality	2.672	2.692	2.658	-0.034
	(0.390)	(0.378)	(0.401)	(0.082)
RM outcomes	0.715	0.706	0.721	0.014
	(0.199)	(0.184)	(0.210)	(0.042)
Monitoring	2.442	2.455	2.433	-0.022
	(0.552)	(0.509)	(0.584)	(0.116)
Evaluation	4.605	4.390	4.754	0.364
	(1.937)	(1.971)	(1.917)	(0.404)
Approved amount millions 2010 US$	85.942	99.290	76.575	-22.715
	(138.626)	(120.228)	(150.534)	(28.649)
N. outcomes over N. components	2.759	2.918	2.651	-0.267
	(1.901)	(2.336)	(1.548)	(0.396)
N. outputs over N. components	5.057	4.710	5.294	0.585
	(4.298)	(2.930)	(5.036)	(0.896)
N. components	3.021	2.875	3.123	0.248
	(1.299)	(1.265)	(1.324)	(0.268)
Observations	97	40	57	97

Table 4. Descriptive Statistics PCRs by CO Year, Execution Performance (PMR)

Variable	(1) Overall Mean	(2) Mean PCR17	(3) Mean PCR18	(4) Diff. PCR18 vs. PCR17
Disbursed share from approved amount	0.920	0.936	0.909	-0.027
	(0.174)	(0.163)	(0.182)	(0.036)
Disbursed under 60% of original amount	0.103	0.075	0.123	0.048
	(0.306)	(0.267)	(0.331)	(0.063)
Disbursed under 50% of original amount	0.052	0.050	0.053	0.003
	(0.222)	(0.221)	(0.225)	(0.046)
N. months between eligibility and approval	11.990	11.425	12.386	0.961
	(8.752)	(8.975)	(8.649)	(1.812)
N. months overrun	12.155	10.575	13.263	2.688
	(11.948)	(10.609)	(12.779)	(2.462)
Last Disbursement Index PMR	0.930	0.904	0.947	0.043^
	(0.131)	(0.148)	(0.117)	(0.029)
Performance Index - Alert/Problem Ratio	0.218	0.234	0.207	-0.027
	(0.246)	(0.236)	(0.254)	(0.055)
Performance Index - Problem Ratio	0.098	0.082	0.109	0.028
	(0.163)	(0.141)	(0.177)	(0.036)
Performance Index - Alert/Problem Ratio 3 first years	0.296	0.314	0.283	-0.030
	(0.321)	(0.306)	(0.334)	(0.072)
Performance Index - Problem Ratio 3 first years	0.133	0.108	0.150	0.042
	(0.234)	(0.196)	(0.257)	(0.052)
Performance Index - Alert/Problem Ratio 3 last years	0.129	0.137	0.123	-0.014
	(0.271)	(0.308)	(0.245)	(0.060)
Performance Index - Problem Ratio 3 last years	0.058	0.059	0.057	-0.002
	(0.175)	(0.209)	(0.149)	(0.039)

Performance Index - Alert/Problem dummy 1st year	0.250	0.235	0.260	0.025
	(0.436)	(0.431)	(0.443)	(0.097)
Performance Index - Problem dummy 1st year	0.143	0.088	0.180	0.092
	(0.352)	(0.288)	(0.388)	(0.078)
Performance Index - Alert/Problem dummy Last year	0.060	0.059	0.060	0.001
	(0.238)	(0.239)	(0.240)	(0.053)
Performance Index - Problem dummy Last year	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Number of Extensions	1.424	1.324	1.490	0.167
	(1.127)	(1.007)	(1.206)	(0.250)
Dummy for at least 1 Extension	0.750	0.794	0.720	-0.074
	(0.436)	(0.410)	(0.454)	(0.097)
Share outputs 1st matrix discontinued	0.191	0.195	0.188	-0.007
	(0.242)	(0.235)	(0.248)	(0.050)
Share outputs last matrix not achieved	0.177	0.142	0.202	0.060
	(0.237)	(0.214)	(0.251)	(0.049)
Cost share from outputs 1st matrix discontinued	0.111	0.088	0.127	0.039
	(0.216)	(0.154)	(0.249)	(0.048)
Cost share from outputs last matrix not achieved	0.221	0.233	0.214	-0.019
	(0.316)	(0.350)	(0.294)	(0.070)
Observations	97	40	57	97

Table 11. Quality at entry characteristics by rating of Success (OVE)

Variable	(1) Overall Mean	(2) < Partly Successful	(3) >= Partly Successful	(4) Diff.
All - Original DEM score	7.783	7.630	7.900	0.270
	(1.381)	(1.454)	(1.325)	(0.286)
S3 - Program Logic	7.902	7.712	8.047	0.334
	(1.455)	(1.457)	(1.451)	(0.301)
S4 - Economic Analysis	7.571	7.032	7.980	0.948^
	(3.268)	(3.725)	(2.841)	(0.673)
S5 - M&E	7.047	6.987	7.092	0.105
	(1.903)	(1.948)	(1.886)	(0.396)
RM quality	2.672	2.588	2.735	0.147*
	(0.390)	(0.434)	(0.344)	(0.080)
RM outcomes	0.715	0.661	0.755	0.094**
	(0.199)	(0.147)	(0.223)	(0.040)
Monitoring	2.442	2.479	2.414	-0.065
	(0.552)	(0.540)	(0.564)	(0.115)
Evaluation	4.605	4.508	4.679	0.171
	(1.937)	(1.982)	(1.917)	(0.403)
Approved amount millions 2010 US$	85.942	75.148	93.845	18.697
	(138.626)	(104.434)	(159.563)	(28.579)
N. outcomes over N. components	2.759	2.826	2.712	-0.114
	(1.901)	(1.922)	(1.902)	(0.395)
N. outputs over N. components	5.057	5.474	4.759	-0.715
	(4.298)	(5.584)	(3.094)	(0.891)
N. components	3.021	3.317	2.804	-0.514*
	(1.299)	(1.404)	(1.182)	(0.263)
Observations	97	41	56	97

Table 12. Execution Performance by rating of Success (OVE)

Variable	(1) Overall Mean	(2) < Partly Successful	(3) >= Partly Successful	(4) Diff.
Disbursed share from approved amount	0.920	0.847	0.973	0.126***
	(0.174)	(0.224)	(0.099)	(0.034)
Disbursed under 60% of original amount	0.103	0.220	0.018	-0.202***
	(0.306)	(0.419)	(0.134)	(0.060)
Disbursed under 50% of original amount	0.052	0.098	0.018	-0.080*
	(0.222)	(0.300)	(0.134)	(0.045)
N. months between eligibility and approval	11.990	12.146	11.875	-0.271
	(8.752)	(9.546)	(8.209)	(1.808)
N. months overrun	12.155	15.415	9.768	-5.647**
	(11.948)	(13.353)	(10.285)	(2.400)
Last Disbursement Index PMR	0.930	0.903	0.951	0.048*
	(0.131)	(0.159)	(0.100)	(0.028)
Performance Index - Alert/Problem Ratio	0.218	0.321	0.137	-0.184***
	(0.246)	(0.282)	(0.178)	(0.050)
Performance Index - Problem Ratio	0.098	0.160	0.049	-0.111***
	(0.163)	(0.193)	(0.115)	(0.034)
Performance Index - Alert/Problem Ratio 3 first years	0.296	0.414	0.202	-0.212***
	(0.321)	(0.346)	(0.269)	(0.067)
Performance Index - Problem Ratio 3 first years	0.133	0.207	0.074	-0.133***
	(0.234)	(0.287)	(0.162)	(0.050)
Performance Index - Alert/Problem Ratio 3 last years	0.129	0.225	0.053	-0.172***
	(0.271)	(0.352)	(0.148)	(0.057)
Performance Index - Problem Ratio 3 last years	0.058	0.117	0.011	-0.106***
	(0.175)	(0.239)	(0.073)	(0.037)
Performance Index - Alert/Problem dummy 1st year	0.250	0.297	0.213	-0.085

	(0.436)	(0.463)	(0.414)	(0.096)
Performance Index - Problem dummy 1st year	0.143	0.162	0.128	-0.035
	(0.352)	(0.374)	(0.337)	(0.078)
Performance Index - Alert/Problem dummy Last year	0.060	0.108	0.021	-0.087*
	(0.238)	(0.315)	(0.146)	(0.052)
Performance Index - Problem dummy Last year	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Number of Extensions	1.424	1.447	1.404	-0.043
	(1.127)	(0.978)	(1.245)	(0.247)
Dummy for at least 1 Extension	0.750	0.811	0.702	-0.109
	(0.436)	(0.397)	(0.462)	(0.096)
Share outputs 1st matrix discontinued	0.191	0.240	0.155	-0.085*
	(0.242)	(0.262)	(0.221)	(0.049)
Share outputs last matrix not achieved	0.177	0.238	0.132	-0.106**
	(0.237)	(0.282)	(0.188)	(0.048)
Cost share from outputs 1st matrix discontinued	0.111	0.144	0.086	-0.058
	(0.216)	(0.263)	(0.171)	(0.047)
Cost share from outputs last matrix not achieved	0.221	0.240	0.207	-0.033
	(0.316)	(0.319)	(0.316)	(0.069)
Observations	97	41	56	97

Table 13. Quality at entry characteristics by Effectiveness rating (OVE)

Variable	(1) Overall Mean	(2) < Satisfactory	(3) >= Satisfactory	(4) Diff.
All - Original DEM score	7.783	7.619	8.177	0.558*
	(1.381)	(1.384)	(1.316)	(0.307)
S3 - Program Logic	7.902	7.819	8.102	0.283
	(1.455)	(1.400)	(1.590)	(0.328)
S4 - Economic Analysis	7.571	7.351	8.096	0.746
	(3.268)	(3.516)	(2.559)	(0.735)
S5 - M&E	7.047	6.762	7.728	0.965**
	(1.903)	(1.936)	(1.665)	(0.419)
RM quality	2.672	2.628	2.776	0.148*
	(0.390)	(0.425)	(0.269)	(0.087)
RM outcomes	0.715	0.679	0.800	0.121***
	(0.199)	(0.180)	(0.219)	(0.043)
Monitoring	2.442	2.463	2.393	-0.070
	(0.552)	(0.570)	(0.511)	(0.125)
Evaluation	4.605	4.300	5.335	1.035**
	(1.937)	(1.910)	(1.835)	(0.425)
Approved amount millions 2010 US$	85.942	84.297	89.800	5.502
	(138.626)	(149.466)	(111.396)	(30.901)
N. outcomes over N. components	2.759	2.721	2.847	0.125
	(1.901)	(1.767)	(2.211)	(0.425)
N. outputs over N. components	5.057	5.155	4.830	-0.325
	(4.298)	(4.721)	(3.170)	(0.960)
N. components	3.021	3.176	2.655	-0.521*
	(1.299)	(1.326)	(1.173)	(0.285)
Observations	97	68	29	97

Table 14. Execution performance indicators by Effectiveness rating (OVE)

Variable	(1) Overall Mean	(2) < Satisfactory	(3) >= Satisfactory	(4) Diff.
Disbursed share from approved amount	0.920	0.891	0.988	0.098**
	(0.174)	(0.201)	(0.033)	(0.038)
Disbursed under 60% of original amount	0.103	0.147	0.000	-0.147**
	(0.306)	(0.357)	(0.000)	(0.066)
Disbursed under 50% of original amount	0.052	0.074	0.000	-0.074^
	(0.222)	(0.263)	(0.000)	(0.049)
N. months between eligibility and approval	11.990	12.412	11.000	-1.412
	(8.752)	(9.353)	(7.201)	(1.946)
N. months overrun	12.155	14.279	7.172	-7.107***
	(11.948)	(12.214)	(9.787)	(2.562)
Last Disbursement Index PMR	0.930	0.924	0.949	0.025
	(0.131)	(0.136)	(0.117)	(0.033)
Performance Index - Alert/Problem Ratio	0.218	0.244	0.140	-0.104*
	(0.246)	(0.255)	(0.200)	(0.061)
Performance Index - Problem Ratio	0.098	0.109	0.064	-0.045
	(0.163)	(0.168)	(0.146)	(0.041)
Performance Index - Alert/Problem Ratio 3 first years	0.296	0.323	0.214	-0.108^
	(0.321)	(0.322)	(0.312)	(0.081)
Performance Index - Problem Ratio 3 first years	0.133	0.148	0.087	-0.061
	(0.234)	(0.245)	(0.195)	(0.059)
Performance Index - Alert/Problem Ratio 3 last years	0.129	0.159	0.040	-0.119*
	(0.271)	(0.298)	(0.128)	(0.067)
Performance Index - Problem Ratio 3 last years	0.058	0.069	0.024	-0.045
	(0.175)	(0.191)	(0.109)	(0.044)
Performance Index - Alert/Problem dummy 1st year	0.250	0.254	0.238	-0.016
	(0.436)	(0.439)	(0.436)	(0.110)
Performance Index - Problem dummy 1st year	0.143	0.127	0.190	0.063
	(0.352)	(0.336)	(0.402)	(0.089)
Performance Index - Alert/Problem dummy Last year	0.060	0.079	0.000	-0.079^
	(0.238)	(0.272)	(0.000)	(0.060)
Performance Index - Problem dummy Last year	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Number of Extensions	1.424	1.453	1.333	-0.120
	(1.127)	(1.022)	(1.426)	(0.285)
Dummy for at least 1 Extension	0.750	0.794	0.619	-0.175^

	(0.436)	(0.408)	(0.498)	(0.109)
Share outputs 1st matrix discontinued	0.191	0.226	0.109	-0.117**
	(0.242)	(0.254)	(0.188)	(0.052)
Share outputs last matrix not achieved	0.177	0.212	0.095	-0.117**
	(0.237)	(0.243)	(0.203)	(0.051)
Cost share from outputs 1st matrix discontinued	0.111	0.119	0.087	-0.032
	(0.216)	(0.221)	(0.201)	(0.054)
Cost share from outputs last matrix not achieved	0.221	0.247	0.148	-0.099
	(0.316)	(0.323)	(0.288)	(0.078)
Observations	97	68	29	97

Table 15. Quality at entry characteristics by rating of Success (MGT)

Variable	(1) Overall Mean	(2) < Partly Successful	(3) >= Partly Successful	(4) Diff.
All - Original DEM score	7.783	7.622	7.844	0.223
	(1.381)	(1.481)	(1.348)	(0.319)
S3 - Program Logic	7.902	7.624	8.007	0.383
	(1.455)	(1.356)	(1.487)	(0.334)
S4 - Economic Analysis	7.571	7.108	7.745	0.637
	(3.268)	(3.749)	(3.079)	(0.753)
S5 - M&E	7.047	7.104	7.025	-0.079
	(1.903)	(1.970)	(1.892)	(0.440)
RM quality	2.672	2.551	2.717	0.166*
	(0.390)	(0.420)	(0.371)	(0.089)
RM outcomes	0.715	0.688	0.725	0.037
	(0.199)	(0.124)	(0.220)	(0.046)
Monitoring	2.442	2.481	2.427	-0.053
	(0.552)	(0.463)	(0.584)	(0.128)
Evaluation	4.605	4.623	4.598	-0.025
	(1.937)	(1.905)	(1.963)	(0.448)
Approved amount millions 2010 US$	85.942	64.167	93.916	29.749
	(138.626)	(106.725)	(148.481)	(31.798)
N. outcomes over N. components	2.759	2.322	2.913	0.591^
	(1.901)	(1.454)	(2.022)	(0.440)
N. outputs over N. components	5.057	5.116	5.036	-0.080
	(4.298)	(6.531)	(3.231)	(1.005)
N. components	3.021	3.308	2.915	-0.392^
	(1.299)	(1.463)	(1.228)	(0.297)
Observations	97	26	71	97

Table 16. Execution Performance by rating of Success (MGT)

Variable	(1) Overall Mean	(2) < Partly Successful	(3) >= Partly Successful	(4) Diff.
Disbursed share from approved amount	0.920	0.816	0.958	0.142***
	(0.174)	(0.252)	(0.116)	(0.037)
Disbursed under 60% of original amount	0.103	0.269	0.042	-0.227***
	(0.306)	(0.452)	(0.203)	(0.066)
Disbursed under 50% of original amount	0.052	0.154	0.014	-0.140***
	(0.222)	(0.368)	(0.119)	(0.049)
N. months between eligibility and approval	11.990	12.346	11.859	-0.487
	(8.752)	(8.980)	(8.728)	(2.016)
N. months overrun	12.155	16.077	10.718	-5.359**
	(11.948)	(14.913)	(10.419)	(2.698)
Last Disbursement Index PMR	0.930	0.879	0.951	0.072**
	(0.131)	(0.176)	(0.103)	(0.031)
Performance Index - Alert/Problem Ratio	0.218	0.381	0.153	-0.228***
	(0.246)	(0.287)	(0.195)	(0.054)
Performance Index - Problem Ratio	0.098	0.199	0.058	-0.141***
	(0.163)	(0.200)	(0.126)	(0.036)
Performance Index - Alert/Problem Ratio 3 first years	0.296	0.458	0.231	-0.228***
	(0.321)	(0.352)	(0.286)	(0.074)
Performance Index - Problem Ratio 3 first years	0.133	0.250	0.086	-0.164***
	(0.234)	(0.299)	(0.185)	(0.054)
Performance Index - Alert/Problem Ratio 3 last years	0.129	0.292	0.064	-0.228***
	(0.271)	(0.372)	(0.184)	(0.061)
Performance Index - Problem Ratio 3 last years	0.058	0.153	0.019	-0.133***
	(0.175)	(0.278)	(0.087)	(0.040)

Performance Index - Alert/Problem dummy 1st year	0.250	0.333	0.217	-0.117
	(0.436)	(0.482)	(0.415)	(0.105)
Performance Index - Problem dummy 1st year	0.143	0.208	0.117	-0.092
	(0.352)	(0.415)	(0.324)	(0.085)
Performance Index - Alert/Problem dummy Last year	0.060	0.125	0.033	-0.092^
	(0.238)	(0.338)	(0.181)	(0.057)
Performance Index - Problem dummy Last year	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Number of Extensions	1.424	1.480	1.400	-0.080
	(1.127)	(1.085)	(1.153)	(0.270)
Dummy for at least 1 Extension	0.750	0.750	0.750	-0.000
	(0.436)	(0.442)	(0.437)	(0.106)
Share outputs 1st matrix discontinued	0.191	0.239	0.173	-0.065
	(0.242)	(0.269)	(0.230)	(0.055)
Share outputs last matrix not achieved	0.177	0.290	0.136	-0.154***
	(0.237)	(0.285)	(0.204)	(0.052)
Cost share from outputs 1st matrix discontinued	0.111	0.145	0.098	-0.046
	(0.216)	(0.292)	(0.180)	(0.053)
Cost share from outputs last matrix not achieved	0.221	0.341	0.172	-0.168**
	(0.316)	(0.348)	(0.291)	(0.073)
Observations	97	26	71	97

Table 17. Quality at entry characteristics by Effectiveness rating (MGT)

Variable	(1) Overall Mean	(2) < Satisfactory	(3) >= Satisfactory	(4) Diff.
All - Original DEM score	7.783	7.640	8.008	0.367
	(1.381)	(1.356)	(1.409)	(0.290)
S3 - Program Logic	7.902	7.849	7.987	0.138
	(1.455)	(1.386)	(1.574)	(0.308)
S4 - Economic Analysis	7.571	7.405	7.830	0.425
	(3.268)	(3.447)	(2.993)	(0.690)
S5 - M&E	7.047	6.855	7.348	0.494
	(1.903)	(1.947)	(1.818)	(0.399)
RM quality	2.672	2.641	2.720	0.080
	(0.390)	(0.405)	(0.365)	(0.082)
RM outcomes	0.715	0.696	0.744	0.049
	(0.199)	(0.160)	(0.248)	(0.042)
Monitoring	2.442	2.464	2.407	-0.057
	(0.552)	(0.545)	(0.567)	(0.117)
Evaluation	4.605	4.390	4.941	0.550^
	(1.937)	(1.948)	(1.898)	(0.406)
Approved amount millions 2010 US$	85.942	97.275	69.088	-28.187
	(138.626)	(167.830)	(76.228)	(28.712)
N. outcomes over N. components	2.759	2.781	2.727	-0.054
	(1.901)	(1.811)	(2.049)	(0.397)
N. outputs over N. components	5.057	5.488	4.427	-1.061
	(4.298)	(4.953)	(3.055)	(0.891)
N. components	3.021	3.172	2.795	-0.378^
	(1.299)	(1.272)	(1.321)	(0.268)
Observations	97	58	39	97

Table 18. Execution performance indicators by Effectiveness rating (MGT)

Variable	(1) Overall Mean	(2) < Satisfactory	(3) >= Satisfactory	(4) Diff.
Disbursed share from approved amount	0.920	0.891	0.963	0.072**
	(0.174)	(0.199)	(0.119)	(0.036)
Disbursed under 60% of original amount	0.103	0.155	0.026	-0.130**
	(0.306)	(0.365)	(0.160)	(0.062)
Disbursed under 50% of original amount	0.052	0.069	0.026	-0.043
	(0.222)	(0.256)	(0.160)	(0.046)
N. months between eligibility and approval	11.990	12.397	11.385	-1.012
	(8.752)	(9.780)	(7.029)	(1.819)
N. months overrun	12.155	14.190	9.128	-5.061**
	(11.948)	(12.361)	(10.759)	(2.432)
Last Disbursement Index PMR	0.930	0.910	0.964	0.054*
	(0.131)	(0.144)	(0.098)	(0.029)
Performance Index - Alert/Problem Ratio	0.218	0.261	0.143	-0.118**
	(0.246)	(0.270)	(0.179)	(0.054)
Performance Index - Problem Ratio	0.098	0.115	0.070	-0.045
	(0.163)	(0.176)	(0.135)	(0.037)
Performance Index - Alert/Problem Ratio 3 first years	0.296	0.333	0.231	-0.102^
	(0.321)	(0.340)	(0.281)	(0.072)
Performance Index - Problem Ratio 3 first years	0.133	0.157	0.091	-0.066
	(0.234)	(0.258)	(0.182)	(0.053)
Performance Index - Alert/Problem Ratio 3 last years	0.129	0.182	0.038	-0.145**
	(0.271)	(0.317)	(0.119)	(0.059)
Performance Index - Problem Ratio 3 last years	0.058	0.075	0.027	-0.049
	(0.175)	(0.203)	(0.106)	(0.039)

Performance Index - Alert/Problem dummy 1st year	0.250	0.226	0.290	0.064
	(0.436)	(0.423)	(0.461)	(0.099)
Performance Index - Problem dummy 1st year	0.143	0.113	0.194	0.080
	(0.352)	(0.320)	(0.402)	(0.080)
Performance Index - Alert/Problem dummy Last year	0.060	0.094	0.000	-0.094*
	(0.238)	(0.295)	(0.000)	(0.053)
Performance Index - Problem dummy Last year	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Number of Extensions	1.424	1.426	1.419	-0.007
	(1.127)	(1.002)	(1.336)	(0.256)
Dummy for at least 1 Extension	0.750	0.792	0.677	-0.115
	(0.436)	(0.409)	(0.475)	(0.098)
Share outputs 1st matrix discontinued	0.191	0.225	0.140	-0.085*
	(0.242)	(0.261)	(0.202)	(0.050)
Share outputs last matrix not achieved	0.177	0.230	0.099	-0.131***
	(0.237)	(0.245)	(0.203)	(0.047)
Cost share from outputs 1st matrix discontinued	0.111	0.118	0.099	-0.020
	(0.216)	(0.228)	(0.197)	(0.049)
Cost share from outputs last matrix not achieved	0.221	0.274	0.133	-0.141**
	(0.316)	(0.339)	(0.253)	(0.069)
Observations	97	58	39	97

Annex A

Notes on PCR performance by Sector

Tables A1 to A4 in the Appendix presents descriptive statistics for PCR17and PCR18 by sector. We present a short summary of key findings and observations in this section, cognizant of the fact that these are based on few observations.

CSD, with 8 and 9 PCRs in 2017 and 2018, respectively, saw an overall improvement from PCR17 to PCR18. Projects rated => partly successful increased by 18%; projects with effectiveness rated =>satisfactory increased by 32%. On average, the DEM for PCR18 were better on project logic, although a bit worse overall. In terms of execution performance, PCR18 experienced higher disbursement from approved amount; cero projects executed less than 60%, compared to 25% for PCR17; and PCR18 experienced about 5 months of overrun. This suggests that improved execution likely played a role in improved success.

IFD, with 17 and 17 PCR in 2017 and 2018, respectively, saw an overall drop from PCR17 to PCR18. Projects rated => partly successful decreased by 18%; projects with effectiveness rated =>satisfactory decreased by 12%. On average, the DEM were better overall for PCR18. In terms of execution performance, PCR18 experienced 10.4% less disbursement from approved amount, 18% of PCR18 executed less than 60%, compared to 0% for PCR17. Finally, PCR18 had slightly higher overrun and were significantly more likely to be in alert or classified as a problem project in its last 3 years of execution. This suggests that that poorer execution performance might have played a role in the decline success of PCR18.

INE, with 9 and 14 PCR in 2017 and 2018, respectively, saw an overall drop from PCR17 to PCR18. Projects rated => partly successful decreased by 40%; projects with effectiveness rated =>satisfactory decreased by 16%. On average, the DEM were better overall for PCR18. In terms of execution performance, PCR18 experienced 6% less disbursement from approved amount, 14% of PCR18 executed less than 60%, compared to 0% for PCR17. Finally, PCR18 had slightly higher overrun (2.7 months). This suggests that that poorer execution performance might have played a role in the decline success of PCR18.

INT only had 1 and 2 PCRs for 2017 and 2018 in our dataset, respectively. None were rated =>Partly Successful.

SCL, with 5 and 15 PCR in 2017 and 2018, respectively, saw an overall drop from PCR17 to PCR18. Projects rated => partly successful decreased by 27%; projects with effectiveness rated =>satisfactory decreased by 40%. On average, the DEM was largely worse overall for PCR18. In particular, results matrix outcomes were rated significantly lower for PCR18. In terms of execution performance, PCR18 experienced 8.7% less disbursement from approved amount, 13% of PCR18 executed less than 60%, compared to 0% for PCR17. Finally, PCR18 had higher overrun (5.7 months). This suggests that that poorer execution performance might have played a role in the decline success of PCR18, as well as the poorer quality at entry.

Appendix A

Table A1. Descriptive statistics for PCR17and PCR18 (CSD)

Variable	(1) Overall Mean	(2) Mean PCR17	(3) Mean PCR18	(4) Diff. PCR18 vs. PCR17
Mngmt. Rating = Highly Successful	0.059	0.000	0.111	0.111
	(0.243)	(0.000)	(0.333)	(0.118)
Mngmt. Rating = Successful	0.294	0.250	0.333	0.083
	(0.470)	(0.463)	(0.500)	(0.235)
Mngmt. Rating = Partly Successful	0.353	0.375	0.333	-0.042
	(0.493)	(0.518)	(0.500)	(0.247)
Mngmt. Rating = Partly Unsuccessful	0.176	0.375	0.000	-0.375**
	(0.393)	(0.518)	(0.000)	(0.172)
Mngmt. Rating = Unsuccessful	0.118	0.000	0.222	0.222^
	(0.332)	(0.000)	(0.441)	(0.156)
Mngmt. Rating = Highly Unsuccessful	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
OVE = Highly Successful	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
OVE = Successful	0.235	0.125	0.333	0.208
	(0.437)	(0.354)	(0.500)	(0.213)
OVE = Partly Successful	0.235	0.250	0.222	-0.028
	(0.437)	(0.463)	(0.441)	(0.219)
OVE = Partly Unsuccessful	0.353	0.625	0.111	-0.514**
	(0.493)	(0.518)	(0.333)	(0.209)
OVE = Unsuccessful	0.118	0.000	0.222	0.222^
	(0.332)	(0.000)	(0.441)	(0.156)
OVE = Highly Unsuccessful	0.059	0.000	0.111	0.111
	(0.243)	(0.000)	(0.333)	(0.118)

Mngmt. Rating >=Partly Successful	0.706	0.625	0.778	0.153
	(0.470)	(0.518)	(0.441)	(0.232)
OVE Rating >=Partly Successful	0.471	0.375	0.556	0.181
	(0.514)	(0.518)	(0.527)	(0.254)
Mngmt. relevance (>=satisfactory)	0.765	0.875	0.667	-0.208
	(0.437)	(0.354)	(0.500)	(0.213)
Mngmt. effectivenes (>=satisfactory)	0.471	0.375	0.556	0.181
	(0.514)	(0.518)	(0.527)	(0.254)
Mngmt. efficiency (>=satisfactory)	0.471	0.375	0.556	0.181
	(0.514)	(0.518)	(0.527)	(0.254)
Mngmt. sustainability (>=satisfactory)	0.706	0.625	0.778	0.153
	(0.470)	(0.518)	(0.441)	(0.232)
OVE relevance (>=satisfactory)	0.824	1.000	0.667	-0.333*
	(0.393)	(0.000)	(0.500)	(0.177)
OVE effectiveness (>=satisfactory)	0.294	0.125	0.444	0.319^
	(0.470)	(0.354)	(0.527)	(0.221)
OVE efficiency (>=satisfactory)	0.353	0.375	0.333	-0.042
	(0.493)	(0.518)	(0.500)	(0.247)
OVE sustainability (>=satisfactory)	0.647	0.750	0.556	-0.194
	(0.493)	(0.463)	(0.527)	(0.242)
Mngmt. relevance score	3.294	3.375	3.222	-0.153
	(0.849)	(0.744)	(0.972)	(0.424)
Mngmt. effectivenes score	2.412	2.375	2.444	0.069
	(0.795)	(0.518)	(1.014)	(0.399)
Mngmt. efficiency score	2.647	2.625	2.667	0.042
	(0.786)	(0.916)	(0.707)	(0.394)
Mngmt. sustainability score	2.765	2.625	2.889	0.264
	(0.562)	(0.518)	(0.601)	(0.274)

OVE relevance score	3.059	3.125	3.000	-0.125
	(0.659)	(0.354)	(0.866)	(0.329)
OVE effectiveness score	2.059	1.875	2.222	0.347
	(0.899)	(0.641)	(1.093)	(0.442)
OVE efficiency score	2.353	2.375	2.333	-0.042
	(0.862)	(0.916)	(0.866)	(0.432)
OVE sustainability score	2.706	2.875	2.556	-0.319
	(0.772)	(0.641)	(0.882)	(0.378)
Changes to PU due to binary rules (%)	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Changes to PS due to binary rules (%)	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Overall succesful downgrade	0.235	0.250	0.222	-0.028
	(0.437)	(0.463)	(0.441)	(0.219)
Relevance satisfactory downgrade	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Effectiveness satisfactory downgrade	0.176	0.250	0.111	-0.139
	(0.393)	(0.463)	(0.333)	(0.194)
Efficiency satisfactory downgrade	0.118	0.000	0.222	0.222^
	(0.332)	(0.000)	(0.441)	(0.156)
Sustainability satisfactory downgrade	0.118	0.000	0.222	0.222^
	(0.332)	(0.000)	(0.441)	(0.156)
Mngmt. PCR Score	2.706	2.675	2.733	0.058
	(0.575)	(0.385)	(0.728)	(0.288)
OVE PCR Score	2.447	2.425	2.467	0.042
	(0.610)	(0.345)	(0.800)	(0.306)
OVE PCR Score - Mngmt. Score	-0.259	-0.250	-0.267	-0.017
	(0.345)	(0.366)	(0.346)	(0.173)

OVE Rating Simulation 1 (Relevance) >=Partly Successful	0.471	0.375	0.556	0.181
	(0.514)	(0.518)	(0.527)	(0.254)
OVE Rating Simulation 2 (Effectiveness) >=Partly Successful	0.588	0.625	0.556	-0.069
	(0.507)	(0.518)	(0.527)	(0.254)
OVE Rating Simulation 3 (Efficiency) >=Partly Successful	0.471	0.375	0.556	0.181
	(0.514)	(0.518)	(0.527)	(0.254)
OVE Rating Simulation 4 (Sustainability) >=Partly Successful	0.471	0.375	0.556	0.181
	(0.514)	(0.518)	(0.527)	(0.254)
OVE Rating Simulation 5 (No Binary Rules) >=Partly Successful	0.471	0.375	0.556	0.181
	(0.514)	(0.518)	(0.527)	(0.254)
OVE Rating Simulation 6 (Equal Weights - No Br) >=Partly Successful	0.588	0.625	0.556	-0.069
	(0.507)	(0.518)	(0.527)	(0.254)
All - Original DEM score	7.915	7.940	7.893	-0.047
	(1.347)	(1.586)	(1.194)	(0.676)
S3 - Program Logic	8.148	7.846	8.416	0.570
	(1.324)	(1.235)	(1.415)	(0.648)
S4 - Economic Analysis	7.388	7.125	7.622	0.497
	(3.757)	(4.518)	(3.199)	(1.881)
S5 - M&E	7.399	7.717	7.115	-0.602
	(1.673)	(1.820)	(1.584)	(0.825)
RM quality	2.821	2.775	2.862	0.087
	(0.200)	(0.201)	(0.201)	(0.098)
RM outcomes	0.652	0.675	0.632	-0.043
	(0.118)	(0.139)	(0.101)	(0.058)
Monitoring	2.882	2.875	2.889	0.014
	(0.218)	(0.231)	(0.220)	(0.110)
Evaluation	4.517	4.843	4.227	-0.616
	(1.810)	(1.909)	(1.778)	(0.894)

Approved amount millions 2010 US$	65.130	86.501	46.133	-40.369
	(68.462)	(82.815)	(50.104)	(32.738)
Disbursed share from approved amount	0.898	0.839	0.951	0.112
	(0.204)	(0.280)	(0.087)	(0.098)
Disbursed under 60% of original amount	0.118	0.250	0.000	-0.250^
	(0.332)	(0.463)	(0.000)	(0.154)
Disbursed under 50% of original amount	0.059	0.125	0.000	-0.125
	(0.243)	(0.354)	(0.000)	(0.117)
N. months between eligibility and approval	14.294	13.875	14.667	0.792
	(8.615)	(7.900)	(9.670)	(4.319)
N. months overrun	16.882	14.000	19.444	5.444
	(11.789)	(12.189)	(11.501)	(5.747)
Last Disbursement Index PMR	0.958	0.964	0.952	-0.012
	(0.071)	(0.044)	(0.091)	(0.035)
Performance Index - Alert Ratio	0.285	0.351	0.225	-0.126
	(0.264)	(0.286)	(0.243)	(0.128)
Performance Index - Problem Ratio	0.109	0.110	0.108	-0.002
	(0.203)	(0.199)	(0.218)	(0.102)
Performance Index - Alert Ratio 3 first years	0.451	0.500	0.407	-0.093
	(0.390)	(0.356)	(0.434)	(0.194)
Performance Index - Problem Ratio 3 first years	0.157	0.125	0.185	0.060
	(0.314)	(0.248)	(0.377)	(0.157)
Performance Index - Alert Ratio 3 last years	0.137	0.208	0.074	-0.134
	(0.265)	(0.354)	(0.147)	(0.128)
Performance Index - Problem Ratio 3 last years	0.078	0.125	0.037	-0.088
	(0.187)	(0.248)	(0.111)	(0.091)
Performance Index - Alert dummy 1st year	0.235	0.250	0.222	-0.028
	(0.437)	(0.463)	(0.441)	(0.219)

Performance Index - Problem dummy Last year	0.176	0.125	0.222	0.097
	(0.393)	(0.354)	(0.441)	(0.196)
Performance Index - Alert dummy Last year	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Performance Index - Problem dummy Last year	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Number of Extensions	1.765	1.750	1.778	0.028
	(1.147)	(1.282)	(1.093)	(0.576)
Dummy for at least 1 Extension	0.882	0.875	0.889	0.014
	(0.332)	(0.354)	(0.333)	(0.167)
Observations	17	8	9	17

Table A2. Descriptive statistics for PCR17and PCR18 (IFD)

Variable	(1) Overall Mean	(2) Mean PCR17	(3) Mean PCR18	(4) Difference PCR18 vs.PCR17
Mngmt. Rating = Highly Successful	0.088	0.059	0.118	0.059
	(0.288)	(0.243)	(0.332)	(0.100)
Mngmt. Rating = Successful	0.176	0.176	0.176	-0.000
	(0.387)	(0.393)	(0.393)	(0.135)
Mngmt. Rating = Partly Successful	0.500	0.588	0.412	-0.176
	(0.508)	(0.507)	(0.507)	(0.174)
Mngmt. Rating = Partly Unsuccessful	0.147	0.059	0.235	0.176^
	(0.359)	(0.243)	(0.437)	(0.121)
Mngmt. Rating = Unsuccessful	0.088	0.118	0.059	-0.059
	(0.288)	(0.332)	(0.243)	(0.100)
Mngmt. Rating = Highly Unsuccessful	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
OVE = Highly Successful	0.029	0.059	0.000	-0.059
	(0.171)	(0.243)	(0.000)	(0.059)
OVE = Successful	0.206	0.235	0.176	-0.059
	(0.410)	(0.437)	(0.393)	(0.143)
OVE = Partly Successful	0.382	0.412	0.353	-0.059
	(0.493)	(0.507)	(0.493)	(0.171)
OVE = Partly Unsuccessful	0.206	0.118	0.294	0.176
	(0.410)	(0.332)	(0.470)	(0.140)
OVE = Unsuccessful	0.176	0.176	0.176	0.000
	(0.387)	(0.393)	(0.393)	(0.135)
OVE = Highly Unsuccessful	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)

Mngmt. Rating >=Partly Successful	0.765	0.824	0.706	-0.118
	(0.431)	(0.393)	(0.470)	(0.149)
OVE Rating >=Partly Successful	0.618	0.706	0.529	-0.176
	(0.493)	(0.470)	(0.514)	(0.169)
Mngmt. relevance (>=satisfactory)	0.912	0.941	0.882	-0.059
	(0.288)	(0.243)	(0.332)	(0.100)
Mngmt. effectivenes (>=satisfactory)	0.353	0.294	0.412	0.118
	(0.485)	(0.470)	(0.507)	(0.168)
Mngmt. efficiency (>=satisfactory)	0.794	0.765	0.824	0.059
	(0.410)	(0.437)	(0.393)	(0.143)
Mngmt. sustainability (>=satisfactory)	0.882	0.882	0.882	0.000
	(0.327)	(0.332)	(0.332)	(0.114)
OVE relevance (>=satisfactory)	0.853	0.882	0.824	-0.059
	(0.359)	(0.332)	(0.393)	(0.125)
OVE effectiveness (>=satisfactory)	0.294	0.353	0.235	-0.118
	(0.462)	(0.493)	(0.437)	(0.160)
OVE efficiency (>=satisfactory)	0.765	0.765	0.765	0.000
	(0.431)	(0.437)	(0.437)	(0.150)
OVE sustainability (>=satisfactory)	0.824	0.765	0.882	0.118
	(0.387)	(0.437)	(0.332)	(0.133)
Mngmt. relevance score	3.471	3.706	3.235	-0.471**
	(0.662)	(0.588)	(0.664)	(0.215)
Mngmt. effectivenes score	2.265	2.294	2.235	-0.059
	(0.931)	(0.985)	(0.903)	(0.324)
Mngmt. efficiency score	3.259	3.308	3.214	-0.093
	(0.944)	(1.109)	(0.802)	(0.370)
Mngmt. sustainability score	3.147	3.059	3.235	0.176
	(0.702)	(0.748)	(0.664)	(0.243)

OVE relevance score	3.029	3.176	2.882	-0.294^
	(0.577)	(0.636)	(0.485)	(0.194)
OVE effectiveness score	2.000	2.235	1.765	-0.471^
	(0.921)	(0.970)	(0.831)	(0.310)
OVE efficiency score	3.037	3.231	2.857	-0.374
	(0.980)	(1.092)	(0.864)	(0.378)
OVE sustainability score	3.000	2.824	3.176	0.353^
	(0.696)	(0.728)	(0.636)	(0.234)
Changes to PU due to binary rules (%)	0.029	0.000	0.059	0.059
	(0.171)	(0.000)	(0.243)	(0.059)
Changes to PS due to binary rules (%)	0.294	0.471	0.118	-0.353**
	(0.462)	(0.514)	(0.332)	(0.149)
Overall succesful downgrade	0.147	0.118	0.176	0.059
	(0.359)	(0.332)	(0.393)	(0.125)
Relevance satisfactory downgrade	0.059	0.059	0.059	0.000
	(0.239)	(0.243)	(0.243)	(0.083)
Effectiveness satisfactory downgrade	0.088	0.000	0.176	0.176*
	(0.288)	(0.000)	(0.393)	(0.095)
Efficiency satisfactory downgrade	0.059	0.000	0.118	0.118^
	(0.239)	(0.000)	(0.332)	(0.081)
Sustainability satisfactory downgrade	0.059	0.118	0.000	-0.118^
	(0.239)	(0.332)	(0.000)	(0.081)
Mngmt. PCR Score	2.835	2.882	2.788	-0.094
	(0.577)	(0.621)	(0.545)	(0.200)
OVE PCR Score	2.582	2.706	2.459	-0.247
	(0.571)	(0.592)	(0.537)	(0.194)
OVE PCR Score - Mngmt. Score	-0.253	-0.176	-0.329	-0.153
	(0.349)	(0.291)	(0.393)	(0.119)

OVE Rating Simulation 1 (Relevance) >=Partly Successful	0.647	0.765	0.529	-0.235^
	(0.485)	(0.437)	(0.514)	(0.164)
OVE Rating Simulation 2 (Effectiveness) >=Partly Successful	0.676	0.765	0.588	-0.176
	(0.475)	(0.437)	(0.507)	(0.162)
OVE Rating Simulation 3 (Efficiency) >=Partly Successful	0.647	0.765	0.529	-0.235^
	(0.485)	(0.437)	(0.514)	(0.164)
OVE Rating Simulation 4 (Sustainability) >=Partly Successful	0.618	0.765	0.471	-0.294*
	(0.493)	(0.437)	(0.514)	(0.164)
OVE Rating Simulation 5 (No Binary Rules) >=Partly Successful	0.618	0.706	0.529	-0.176
	(0.493)	(0.470)	(0.514)	(0.169)
OVE Rating Simulation 6 (Equal Weights - No Br) >=Partly Successful	0.735	0.824	0.647	-0.176
	(0.448)	(0.393)	(0.493)	(0.153)
All - Original DEM score	7.516	7.315	7.705	0.389
	(1.301)	(1.348)	(1.266)	(0.455)
S3 - Program Logic	7.653	7.432	7.860	0.428
	(1.267)	(1.363)	(1.173)	(0.442)
S4 - Economic Analysis	7.033	6.844	7.212	0.368
	(3.564)	(3.613)	(3.619)	(1.260)
S5 - M&E	6.745	6.464	7.009	0.545
	(1.594)	(1.543)	(1.641)	(0.555)
RM quality	2.681	2.657	2.702	0.045
	(0.388)	(0.479)	(0.291)	(0.137)
RM outcomes	0.743	0.712	0.774	0.062
	(0.214)	(0.231)	(0.199)	(0.075)
Monitoring	2.238	2.187	2.285	0.098
	(0.592)	(0.581)	(0.617)	(0.209)
Evaluation	4.507	4.276	4.724	0.448
	(1.685)	(1.853)	(1.535)	(0.591)

Approved amount millions 2010 US$	111.114	99.347	122.881	23.534
	(192.041)	(134.253)	(240.305)	(66.761)
Disbursed share from approved amount	0.915	0.967	0.863	-0.104*
	(0.170)	(0.058)	(0.225)	(0.056)
Disbursed under 60% of original amount	0.088	0.000	0.176	0.176*
	(0.288)	(0.000)	(0.393)	(0.095)
Disbursed under 50% of original amount	0.059	0.000	0.118	0.118^
	(0.239)	(0.000)	(0.332)	(0.081)
N. months between eligibility and approval	12.088	10.647	13.529	2.882
	(10.495)	(11.694)	(9.274)	(3.620)
N. months overrun	8.794	8.412	9.176	0.765
	(10.645)	(10.530)	(11.069)	(3.705)
Last Disbursement Index PMR	0.899	0.891	0.906	0.015
	(0.147)	(0.151)	(0.150)	(0.059)
Performance Index - Alert Ratio	0.244	0.232	0.255	0.023
	(0.244)	(0.171)	(0.300)	(0.098)
Performance Index - Problem Ratio	0.095	0.065	0.120	0.054
	(0.161)	(0.119)	(0.190)	(0.064)
Performance Index - Alert Ratio 3 first years	0.288	0.333	0.250	-0.083
	(0.265)	(0.284)	(0.251)	(0.105)
Performance Index - Problem Ratio 3 first years	0.109	0.111	0.107	-0.004
	(0.216)	(0.217)	(0.223)	(0.087)
Performance Index - Alert Ratio 3 last years	0.147	0.056	0.226	0.171^
	(0.276)	(0.130)	(0.343)	(0.105)
Performance Index - Problem Ratio 3 last years	0.058	0.000	0.107	0.107*
	(0.141)	(0.000)	(0.180)	(0.052)
Performance Index - Alert dummy 1st year	0.269	0.333	0.214	-0.119
	(0.452)	(0.492)	(0.426)	(0.180)

Performance Index - Problem dummy Last year	0.077	0.083	0.071	-0.012
	(0.272)	(0.289)	(0.267)	(0.109)
Performance Index - Alert dummy Last year	0.115	0.083	0.143	0.060
	(0.326)	(0.289)	(0.363)	(0.130)
Performance Index - Problem dummy Last year	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Number of Extensions	0.778	0.833	0.733	-0.100
	(0.698)	(0.577)	(0.799)	(0.275)
Dummy for at least 1 Extension	0.615	0.750	0.500	-0.250
	(0.496)	(0.452)	(0.519)	(0.193)
Observations	34	17	17	34

Table A3. Descriptive statistics for PCR17and PCR18 (INE)

Variable	(1) Overall Mean	(2) Mean PCR17	(3) Mean PCR18	(4) Diff. PCR18 vs.PCR17
Mngmt. Rating = Highly Successful	0.043	0.000	0.071	0.071
	(0.209)	(0.000)	(0.267)	(0.090)
Mngmt. Rating = Successful	0.435	0.556	0.357	-0.198
	(0.507)	(0.527)	(0.497)	(0.217)
Mngmt. Rating = Partly Successful	0.304	0.444	0.214	-0.230
	(0.470)	(0.527)	(0.426)	(0.200)
Mngmt. Rating = Partly Unsuccessful	0.217	0.000	0.357	0.357**
	(0.422)	(0.000)	(0.497)	(0.167)
Mngmt. Rating = Unsuccessful	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Mngmt. Rating = Highly Unsuccessful	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
OVE = Highly Successful	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
OVE = Successful	0.348	0.444	0.286	-0.159
	(0.487)	(0.527)	(0.469)	(0.210)
OVE = Partly Successful	0.304	0.444	0.214	-0.230
	(0.470)	(0.527)	(0.426)	(0.200)
OVE = Partly Unsuccessful	0.261	0.111	0.357	0.246
	(0.449)	(0.333)	(0.497)	(0.189)
OVE = Unsuccessful	0.043	0.000	0.071	0.071
	(0.209)	(0.000)	(0.267)	(0.090)
OVE = Highly Unsuccessful	0.043	0.000	0.071	0.071
	(0.209)	(0.000)	(0.267)	(0.090)

Mngmt. Rating >=Partly Successful	0.783	1.000	0.643	-0.357**
	(0.422)	(0.000)	(0.497)	(0.167)
OVE Rating >=Partly Successful	0.652	0.889	0.500	-0.389*
	(0.487)	(0.333)	(0.519)	(0.195)
Mngmt. relevance (>=satisfactory)	1.000	1.000	1.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Mngmt. effectivenes (>=satisfactory)	0.522	0.556	0.500	-0.056
	(0.511)	(0.527)	(0.519)	(0.223)
Mngmt. efficiency (>=satisfactory)	0.739	0.667	0.786	0.119
	(0.449)	(0.500)	(0.426)	(0.195)
Mngmt. sustainability (>=satisfactory)	0.870	0.889	0.857	-0.032
	(0.344)	(0.333)	(0.363)	(0.150)
OVE relevance (>=satisfactory)	0.957	1.000	0.929	-0.071
	(0.209)	(0.000)	(0.267)	(0.090)
OVE effectiveness (>=satisfactory)	0.348	0.444	0.286	-0.159
	(0.487)	(0.527)	(0.469)	(0.210)
OVE efficiency (>=satisfactory)	0.652	0.556	0.714	0.159
	(0.487)	(0.527)	(0.469)	(0.210)
OVE sustainability (>=satisfactory)	0.783	0.889	0.714	-0.175
	(0.422)	(0.333)	(0.469)	(0.180)
Mngmt. relevance score	3.652	3.889	3.500	-0.389*
	(0.487)	(0.333)	(0.519)	(0.195)
Mngmt. effectivenes score	2.391	2.556	2.286	-0.270
	(0.722)	(0.527)	(0.825)	(0.310)
Mngmt. efficiency score	3.105	3.125	3.091	-0.034
	(0.875)	(0.991)	(0.831)	(0.418)
Mngmt. sustainability score	3.217	3.444	3.071	-0.373
	(0.671)	(0.726)	(0.616)	(0.282)

OVE relevance score	3.522	3.889	3.286	-0.603**
	(0.593)	(0.333)	(0.611)	(0.223)
OVE effectiveness score	2.087	2.333	1.929	-0.405
	(0.793)	(0.707)	(0.829)	(0.335)
OVE efficiency score	2.737	2.875	2.636	-0.239
	(1.098)	(0.991)	(1.206)	(0.522)
OVE sustainability score	2.870	3.111	2.714	-0.397^
	(0.694)	(0.601)	(0.726)	(0.291)
Changes to PU due to binary rules (%)	0.087	0.000	0.143	0.143
	(0.288)	(0.000)	(0.363)	(0.122)
Changes to PS due to binary rules (%)	0.087	0.222	0.000	-0.222*
	(0.288)	(0.441)	(0.000)	(0.116)
Overall succesful downgrade	0.130	0.111	0.143	0.032
	(0.344)	(0.333)	(0.363)	(0.150)
Relevance satisfactory downgrade	0.043	0.000	0.071	0.071
	(0.209)	(0.000)	(0.267)	(0.090)
Effectiveness satisfactory downgrade	0.174	0.111	0.214	0.103
	(0.388)	(0.333)	(0.426)	(0.168)
Efficiency satisfactory downgrade	0.130	0.111	0.143	0.032
	(0.344)	(0.333)	(0.363)	(0.150)
Sustainability satisfactory downgrade	0.087	0.000	0.143	0.143
	(0.288)	(0.000)	(0.363)	(0.122)
Mngmt. PCR Score	2.948	3.111	2.843	-0.268^
	(0.392)	(0.267)	(0.431)	(0.161)
OVE PCR Score	2.661	2.911	2.500	-0.411*
	(0.570)	(0.362)	(0.631)	(0.233)
OVE PCR Score - Mngmt. Score	-0.287	-0.200	-0.343	-0.143
	(0.340)	(0.245)	(0.388)	(0.145)

OVE Rating Simulation 1 (Relevance) >=Partly Successful	0.609	0.889	0.429	-0.460**
	(0.499)	(0.333)	(0.514)	(0.194)
OVE Rating Simulation 2 (Effectiveness) >=Partly Successful	0.783	1.000	0.643	-0.357**
	(0.422)	(0.000)	(0.497)	(0.167)
OVE Rating Simulation 3 (Efficiency) >=Partly Successful	0.652	0.889	0.500	-0.389*
	(0.487)	(0.333)	(0.519)	(0.195)
OVE Rating Simulation 4 (Sustainability) >=Partly Successful	0.652	0.889	0.500	-0.389*
	(0.487)	(0.333)	(0.519)	(0.195)
OVE Rating Simulation 5 (No Binary Rules) >=Partly Successful	0.652	0.889	0.500	-0.389*
	(0.487)	(0.333)	(0.519)	(0.195)
OVE Rating Simulation 6 (Equal Weights - No Br) >=Partly Successful	0.826	1.000	0.714	-0.286*
	(0.388)	(0.000)	(0.469)	(0.158)
All - Original DEM score	7.520	7.305	7.670	0.365
	(1.448)	(1.413)	(1.510)	(0.638)
S3 - Program Logic	7.764	7.672	7.828	0.156
	(1.743)	(1.622)	(1.884)	(0.773)
S4 - Economic Analysis	8.364	8.500	8.269	-0.231
	(2.300)	(1.500)	(2.781)	(1.021)
S5 - M&E	6.114	5.777	6.348	0.571
	(2.045)	(2.196)	(1.990)	(0.900)
RM quality	2.571	2.624	2.535	-0.089
	(0.490)	(0.407)	(0.553)	(0.217)
RM outcomes	0.731	0.654	0.785	0.130*
	(0.182)	(0.141)	(0.193)	(0.075)
Monitoring	2.253	2.472	2.102	-0.370^
	(0.559)	(0.441)	(0.597)	(0.234)
Evaluation	3.861	3.305	4.246	0.941
	(2.006)	(2.082)	(1.939)	(0.866)

Approved amount millions 2010 US$	79.884	74.365	83.432	9.067
	(116.879)	(104.007)	(128.166)	(51.073)
Disbursed share from approved amount	0.955	0.990	0.933	-0.057
	(0.128)	(0.020)	(0.161)	(0.055)
Disbursed under 60% of original amount	0.087	0.000	0.143	0.143
	(0.288)	(0.000)	(0.363)	(0.122)
Disbursed under 50% of original amount	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
N. months between eligibility and approval	10.870	11.889	10.214	-1.675
	(6.497)	(7.253)	(6.154)	(2.817)
N. months overrun	11.870	10.222	12.929	2.706
	(12.005)	(8.913)	(13.853)	(5.216)
Last Disbursement Index PMR	0.954	0.952	0.955	0.003
	(0.100)	(0.056)	(0.126)	(0.048)
Performance Index - Alert Ratio	0.180	0.205	0.162	-0.043
	(0.167)	(0.213)	(0.133)	(0.079)
Performance Index - Problem Ratio	0.091	0.085	0.096	0.012
	(0.113)	(0.092)	(0.130)	(0.054)
Performance Index - Alert Ratio 3 first years	0.281	0.292	0.273	-0.019
	(0.255)	(0.278)	(0.250)	(0.122)
Performance Index - Problem Ratio 3 first years	0.175	0.167	0.182	0.015
	(0.204)	(0.178)	(0.229)	(0.097)
Performance Index - Alert Ratio 3 last years	0.088	0.125	0.061	-0.064
	(0.187)	(0.248)	(0.135)	(0.088)
Performance Index - Problem Ratio 3 last years	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Performance Index - Alert dummy 1st year	0.263	0.125	0.364	0.239
	(0.452)	(0.354)	(0.505)	(0.208)

Performance Index - Problem dummy Last year	0.211	0.125	0.273	0.148
	(0.419)	(0.354)	(0.467)	(0.197)
Performance Index - Alert dummy Last year	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Performance Index - Problem dummy Last year	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Number of Extensions	1.789	1.500	2.000	0.500
	(1.357)	(1.069)	(1.549)	(0.638)
Dummy for at least 1 Extension	0.737	0.750	0.727	-0.023
	(0.452)	(0.463)	(0.467)	(0.216)
Observations	23	9	14	23

Table A4. Descriptive statistics for PCR17and PCR18 (SCL)

Variable	(1) Overall Mean	(2) Mean PCR17	(3) Mean PCR18	(4) Diff. PCR18 vs. PCR17
Mngmt. Rating = Highly Successful	0.200	0.400	0.133	-0.267
	(0.410)	(0.548)	(0.352)	(0.208)
Mngmt. Rating = Successful	0.150	0.000	0.200	0.200
	(0.366)	(0.000)	(0.414)	(0.189)
Mngmt. Rating = Partly Successful	0.300	0.400	0.267	-0.133
	(0.470)	(0.548)	(0.458)	(0.247)
Mngmt. Rating = Partly Unsuccessful	0.250	0.200	0.267	0.067
	(0.444)	(0.447)	(0.458)	(0.235)
Mngmt. Rating = Unsuccessful	0.100	0.000	0.133	0.133
	(0.308)	(0.000)	(0.352)	(0.160)
Mngmt. Rating = Highly Unsuccessful	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
OVE = Highly Successful	0.150	0.200	0.133	-0.067
	(0.366)	(0.447)	(0.352)	(0.194)
OVE = Successful	0.150	0.400	0.067	-0.333*
	(0.366)	(0.548)	(0.258)	(0.178)
OVE = Partly Successful	0.300	0.200	0.333	0.133
	(0.470)	(0.447)	(0.488)	(0.247)
OVE = Partly Unsuccessful	0.300	0.200	0.333	0.133
	(0.470)	(0.447)	(0.488)	(0.247)
OVE = Unsuccessful	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
OVE = Highly Unsuccessful	0.100	0.000	0.133	0.133
	(0.308)	(0.000)	(0.352)	(0.160)

Mngmt. Rating >=Partly Successful	0.650	0.800	0.600	-0.200
	(0.489)	(0.447)	(0.507)	(0.255)
OVE Rating >=Partly Successful	0.600	0.800	0.533	-0.267
	(0.503)	(0.447)	(0.516)	(0.259)
Mngmt. relevance (>=satisfactory)	1.000	1.000	1.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Mngmt. effectivenes (>=satisfactory)	0.350	0.400	0.333	-0.067
	(0.489)	(0.548)	(0.488)	(0.259)
Mngmt. efficiency (>=satisfactory)	0.800	0.800	0.800	0.000
	(0.410)	(0.447)	(0.414)	(0.218)
Mngmt. sustainability (>=satisfactory)	0.900	1.000	0.867	-0.133
	(0.308)	(0.000)	(0.352)	(0.160)
OVE relevance (>=satisfactory)	0.950	1.000	0.933	-0.067
	(0.224)	(0.000)	(0.258)	(0.118)
OVE effectiveness (>=satisfactory)	0.300	0.600	0.200	-0.400^
	(0.470)	(0.548)	(0.414)	(0.231)
OVE efficiency (>=satisfactory)	0.650	0.400	0.733	0.333^
	(0.489)	(0.548)	(0.458)	(0.247)
OVE sustainability (>=satisfactory)	0.850	1.000	0.800	-0.200
	(0.366)	(0.000)	(0.414)	(0.189)
Mngmt. relevance score	3.700	3.800	3.667	-0.133
	(0.470)	(0.447)	(0.488)	(0.247)
Mngmt. effectivenes score	2.150	2.600	2.000	-0.600
	(0.988)	(0.894)	(1.000)	(0.505)
Mngmt. efficiency score	3.421	3.600	3.357	-0.243
	(0.961)	(0.894)	(1.008)	(0.512)
Mngmt. sustainability score	3.300	3.600	3.200	-0.400
	(0.801)	(0.548)	(0.862)	(0.415)

OVE relevance score	3.500	3.600	3.467	-0.133
	(0.607)	(0.548)	(0.640)	(0.320)
OVE effectiveness score	2.100	2.600	1.933	-0.667^
	(0.968)	(0.548)	(1.033)	(0.489)
OVE efficiency score	2.632	2.600	2.643	0.043
	(0.831)	(0.894)	(0.842)	(0.445)
OVE sustainability score	2.950	3.200	2.867	-0.333
	(0.686)	(0.447)	(0.743)	(0.356)
Changes to PU due to binary rules (%)	0.150	0.000	0.200	0.200
	(0.366)	(0.000)	(0.414)	(0.189)
Changes to PS due to binary rules (%)	0.200	0.400	0.133	-0.267
	(0.410)	(0.548)	(0.352)	(0.208)
Overall succesful downgrade	0.050	0.000	0.067	0.067
	(0.224)	(0.000)	(0.258)	(0.118)
Relevance satisfactory downgrade	0.050	0.000	0.067	0.067
	(0.224)	(0.000)	(0.258)	(0.118)
Effectiveness satisfactory downgrade	0.100	0.000	0.133	0.133
	(0.308)	(0.000)	(0.352)	(0.160)
Efficiency satisfactory downgrade	0.150	0.400	0.067	-0.333*
	(0.366)	(0.548)	(0.258)	(0.178)
Sustainability satisfactory downgrade	0.050	0.000	0.067	0.067
	(0.224)	(0.000)	(0.258)	(0.118)
Mngmt. PCR Score	2.930	3.240	2.827	-0.413
	(0.656)	(0.607)	(0.658)	(0.334)
OVE PCR Score	2.650	2.920	2.560	-0.360
	(0.665)	(0.460)	(0.710)	(0.342)
OVE PCR Score - Mngmt. Score	-0.280	-0.320	-0.267	0.053
	(0.391)	(0.540)	(0.352)	(0.207)

OVE Rating Simulation 1 (Relevance) >=Partly Successful	0.650	0.800	0.600	-0.200
	(0.489)	(0.447)	(0.507)	(0.255)
OVE Rating Simulation 2 (Effectiveness) >=Partly Successful	0.600	0.800	0.533	-0.267
	(0.503)	(0.447)	(0.516)	(0.259)
OVE Rating Simulation 3 (Efficiency) >=Partly Successful	0.600	0.800	0.533	-0.267
	(0.503)	(0.447)	(0.516)	(0.259)
OVE Rating Simulation 4 (Sustainability) >=Partly Successful	0.600	0.800	0.533	-0.267
	(0.503)	(0.447)	(0.516)	(0.259)
OVE Rating Simulation 5 (No Binary Rules) >=Partly Successful	0.600	0.800	0.533	-0.267
	(0.503)	(0.447)	(0.516)	(0.259)
OVE Rating Simulation 6 (Equal Weights - No Br) >=Partly Successful	0.850	1.000	0.800	-0.200
	(0.366)	(0.000)	(0.414)	(0.189)
All - Original DEM score	8.606	9.309	8.372	-0.937^
	(1.206)	(0.554)	(1.285)	(0.601)
S3 - Program Logic	8.235	9.059	7.960	-1.099^
	(1.617)	(0.779)	(1.746)	(0.818)
S4 - Economic Analysis	8.175	9.700	7.667	-2.033^
	(3.001)	(0.671)	(3.314)	(1.518)
S5 - M&E	8.622	8.803	8.562	-0.241
	(1.295)	(0.809)	(1.440)	(0.685)
RM quality	2.611	2.779	2.556	-0.223
	(0.396)	(0.225)	(0.430)	(0.203)
RM outcomes	0.719	0.855	0.674	-0.181^
	(0.250)	(0.101)	(0.270)	(0.126)
Monitoring	2.529	2.500	2.538	0.038
	(0.430)	(0.000)	(0.501)	(0.228)
Evaluation	6.093	6.303	6.023	-0.280
	(1.422)	(0.809)	(1.593)	(0.752)

Approved amount millions 2010 US$	79.760	183.118	45.307	-137.811***
	(103.289)	(150.026)	(54.416)	(44.136)
Disbursed share from approved amount	0.929	0.994	0.908	-0.087
	(0.180)	(0.013)	(0.205)	(0.093)
Disbursed under 60% of original amount	0.100	0.000	0.133	0.133
	(0.308)	(0.000)	(0.352)	(0.160)
Disbursed under 50% of original amount	0.050	0.000	0.067	0.067
	(0.224)	(0.000)	(0.258)	(0.118)
N. months between eligibility and approval	10.250	9.600	10.467	0.867
	(7.840)	(2.074)	(9.054)	(4.154)
N. months overrun	12.300	8.000	13.733	5.733
	(12.844)	(4.637)	(14.454)	(6.679)
Last Disbursement Index PMR	0.948	0.880	0.972	0.092*
	(0.104)	(0.113)	(0.093)	(0.051)
Performance Index - Alert Ratio	0.120	0.000	0.162	0.162
	(0.266)	(0.000)	(0.301)	(0.137)
Performance Index - Problem Ratio	0.075	0.000	0.102	0.102
	(0.163)	(0.000)	(0.184)	(0.084)
Performance Index - Alert Ratio 3 first years	0.140	0.000	0.190	0.190
	(0.320)	(0.000)	(0.363)	(0.165)
Performance Index - Problem Ratio 3 first years	0.088	0.000	0.119	0.119
	(0.187)	(0.000)	(0.211)	(0.096)
Performance Index - Alert Ratio 3 last years	0.088	0.000	0.119	0.119
	(0.218)	(0.000)	(0.248)	(0.113)
Performance Index - Problem Ratio 3 last years	0.053	0.000	0.071	0.071
	(0.167)	(0.000)	(0.193)	(0.088)
Performance Index - Alert dummy 1st year	0.105	0.000	0.143	0.143
	(0.315)	(0.000)	(0.363)	(0.165)

Performance Index - Problem dummy Last year	0.105	0.000	0.143	0.143
	(0.315)	(0.000)	(0.363)	(0.165)
Performance Index - Alert dummy Last year	0.053	0.000	0.071	0.071
	(0.229)	(0.000)	(0.267)	(0.122)
Performance Index - Problem dummy Last year	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Number of Extensions	1.474	1.200	1.571	0.371
	(1.020)	(0.837)	(1.089)	(0.539)
Dummy for at least 1 Extension	0.789	0.800	0.786	-0.014
	(0.419)	(0.447)	(0.426)	(0.225)
Observations	20	5	15	20

Table A5. Achievement Ratio of the Result Indicators – percentages

	All PCRs	Country Departments				Sectors				
		CAN	CCB	CID	CSC	CSD	IFD	INE	INT	SCL
All result indicators										
100%	48.0%	47.5%	31.7%	48.6%	52.6%	59.1%	42.2%	48.1%	63.2%	41.0%
[80%-100%)	6.6%	6.3%	10.0%	3.3%	8.8%	7.8%	9.8%	6.7%	5.3%	3.5%
[60%-80%)	6.8%	7.5%	6.7%	5.0%	8.2%	6.1%	6.9%	9.6%	5.3%	4.9%
[40%-60%)	4.7%	6.3%	6.7%	1.7%	6.2%	2.6%	3.9%	6.7%	5.3%	4.9%
[20%-40%)	3.9%	1.3%	6.7%	5.0%	3.1%	4.3%	3.9%	2.2%	5.3%	4.9%
(0%-20%)	2.9%		3.3%	2.8%	4.1%	3.5%	2.0%	1.5%	5.3%	4.2%
0%	13.6%	17.5%	11.7%	15.5%	10.8%	12.2%	19.6%	13.3%	10.5%	11.1%
No information - Cancelled Outputs	3.7%	0.0%	20.0%	2.8%	1.0%	1.7%				11.8%
No information - Monitoring	9.9%	13.8%	3.3%	15.5%	5.2%	2.6%	11.8%	11.9%		13.9%
Result indicators (Satisfactory+ Specific Objectives)	**36.1%**	**37.5%**	**20.0%**	**28.2%**	**47.9%**	**52.2%**	**36.3%**	**42.2%**	**26.3%**	**18.8%**
100%	78.0%	66.7%	58.3%	88.2%	78.5%	83.3%	75.7%	66.7%	100.0%	88.9%
[80%-100%)	8.6%	13.3%	25.0%	3.9%	7.5%	5.0%	10.8%	14.0%		3.7%
[60%-80%)	7.5%	10.0%	8.3%	3.9%	8.6%	5.0%	2.7%	15.8%		3.7%
[40%-60%)	2.2%	6.7%	8.3%		1.1%	1.7%	2.7%	3.5%		
[20%-40%)	0.5%				1.1%	1.7%				
(0%-20%)	0.5%				1.1%	1.7%				
0%	2.7%	3.3%		3.9%	2.2%	1.7%	8.1%			3.7%
No information										
Result indicators (Partly Unsatisfactory- Specific Objectives)	**63.9%**	**62.5%**	**80.0%**	**71.8%**	**52.1%**	**47.8%**	**63.7%**	**57.8%**	**73.7%**	**81.3%**
100%	31.0%	36.0%	25.0%	33.1%	28.7%	32.7%	23.1%	34.6%	50.0%	29.9%
[80%-100%)	5.5%	2.0%	6.3%	3.1%	9.9%	10.9%	9.2%	1.3%	7.1%	3.4%
[60%-80%)	6.4%	6.0%	6.3%	5.4%	7.9%	7.3%	9.2%	5.1%	7.1%	5.1%
[40%-60%)	6.1%	6.0%	6.3%	2.3%	10.9%	3.6%	4.6%	9.0%	7.1%	6.0%
[20%-40%)	5.8%	2.0%	8.3%	6.9%	5.0%	7.3%	6.2%	3.8%	7.1%	6.0%
(0%-20%)	4.3%		4.2%	3.8%	6.9%	5.5%	3.1%	2.6%	7.1%	5.1%
0%	19.8%	26.0%	14.6%	20.0%	18.8%	23.6%	26.2%	23.1%	14.3%	12.8%
No information - Cancelled Outputs	5.8%		25.0%	3.8%	2.0%	3.6%				14.5%
No information - Monitoring	15.5%	22.0%	4.2%	21.5%	9.9%	5.5%	18.5%	20.5%		17.1%