



Corporate Evaluation

2012 Evaluability

Review of Bank Projects





This work is distributed under a Creative Commons license (CC BY-NC-ND 3.0). You are free to copy, distribute and transmit this work to third-parties, under the following conditions:



Attribution - You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).



Non-Commercial - You may not use this work for commercial purposes.



No Derivative Works - You may not alter, transform, or build upon this work.

Waiver - Any of the above conditions can be waived if you get permission from the copyright holder.

© **Inter-American Development Bank, 2013**

Office of Evaluation and Oversight

1350 New York Avenue, N.W.

Washington, D.C. 20577

www.iadb.org/evaluation

TABLE OF CONTENTS

ACRONYMS AND ABBREVIATIONS

EXECUTIVE SUMMARY

I.	INTRODUCTION	1
II.	EVALUABILITY OF SOVEREIGN GUARANTEED PROJECTS IN 2012	2
	A. The Sample and Method of Analysis.....	2
	B. OVE’s Findings on Evaluability.....	3
	1. Project Logic	6
	2. Economic Analysis	8
	3. Monitoring and Evaluation	9
	C. Classification of Project Evaluability	10
III.	PROJECT MONITORING AFTER THE DEM	11
	A. Sample and Method of Analysis.....	12
	B. Main Findings	12
IV.	CONCLUSIONS AND RECOMMENDATIONS	17
ANNEX A:	<u>PROJECTS SELECTED FOR VALIDATION</u>	

ACRONYMS AND ABBREVIATIONS

AOP	Annual Operation Plan
DEF	Development Effectiveness Framework
DEM	Development Effectiveness Matrix
DEO	Development Effectiveness Overview
IDB-9	Ninth General Capital Increase of the IDB
KNL	Knowledge and Learning Sector
LD	Loan Document
NSG	Non-sovereign-guaranteed
OVE	Office of Evaluation and Oversight
PEP	Project Execution Plan
PMR	Progress Monitoring Report
POD	Proposal for Operation Development
SG	Sovereign-guaranteed
SPD	Office of Strategic Planning and Development Effectiveness

EXECUTIVE SUMMARY

The importance of evaluation as a tool to promote learning and improve effectiveness has been extensively recognized in the Bank and reflected in the attention given to evaluability. As a consequence, OVE is required by the IDB-9 Agreement to report annually on the evaluability of the Bank's portfolios. This is the second annual report and assesses only Sovereign Guaranteed (SG) projects approved in 2012, since the tools for analyzing Non-Sovereign Guaranteed projects have been under revision. In addition to reporting on the overall evaluability of the SG projects approved in 2012, the report also analyzes whether the introduction of the Development Effectiveness Matrix (DEM) in 2011 -- and the strength of DEMs since then -- appears to be linked to stronger project monitoring during implementation through the Progress Monitoring Report (PMR).

In preparing this report OVE reviewed 41 projects approved in 2012, which makes up about one-third of the portfolio approved that year. Their respective DEMs were validated following the guidelines produced by Management in June 2012. In addition OVE reviewed a sample of PMRs for 80 projects approved before December 2011 and currently in execution.

As in the previous year's report, the most common evaluability problems found by OVE were related to project design and logic. OVE's validations led to some lowering of Management's scores, particularly in this section. Despite the problems of evaluability identified in many projects, almost all projects continue to be classified, according to Management's classification system, as "highly evaluable", indicating that the way projects are currently labeled is not very informative and needs to be revised.

With regard to project monitoring, the report finds that projects approved after the implementation of the DEM have slightly better tracking of outputs (though not outcomes), but it does not find a significant connection between better DEMs at entry and better PMRs during implementation. For a number of reasons -- whether because of problems in the original plan or because of changes in the project itself -- project teams often change project monitoring plans after project approval, diverging from the original project DEM either positively or negatively. In addition, some PMRs for projects with imperfect monitoring plans at design have been improved by project teams during implementation, with SPD support.

The report offers the following five recommendations for Management:

- 1. Focus greater attention on ensuring proper problem diagnosis at the project design stage.** As in the 2011 review, incomplete problem diagnosis and inadequate project logic are the main weaknesses in the evaluability of SG projects at entry, and these can make it very difficult to be able to measure actual achievements at completion.
- 2. Revise the classification system for projects' evaluability** (as reported, among other places, in the Development Effectiveness Overview - DEO). The current classification is non-informative, since almost all projects fall in the same (and somewhat misleading) category of "highly evaluable". A more

elaborated classification should be considered, which applies narrower ranges and more accurate labels for each category.

3. **Strengthen SPD's evaluability notes to better explain the DEM scores.** The evaluability notes should provide more information regarding the evaluability issues identified in the project and be given more weight in the overall assessment of evaluability.
4. **Increase integration between the documents produced during project design (POD, AOP, PEP, and Results Matrix) and the PMR, and provide adequate guidance and training to ensure their proper preparation by project teams.** Teams need clear guidance in preparing project documents at both the design and monitoring stages in order to improve their overall quality and minimize the need for changes over time.
5. **Ensure that the information from the approved Results Matrix is entered in the system prior to implementation and any changes during implementation are tracked.** Preparation of PMRs at the design stage can help teams achieve better DEMs and make changes in indicators after approval more visible. Ensuring that any changes and the need for them are well documented from the beginning can enhance accountability and provide information for future projects.

I. INTRODUCTION

- 1.1 **The importance of evaluation as a tool to promote learning and improve effectiveness has been extensively recognized in the Bank and reflected in the attention given to evaluability.** Since 2009 Management has worked to improve projects' evaluability and to improve the Development Effectiveness Matrix (DEM) to better assess the evaluability of the Bank's portfolio. The Bank will be able to report on results only if projects are both designed to be evaluable and properly monitored throughout implementation.
- 1.2 **Following the requirements of the IDB-9 agreement, the Office of Evaluation and Oversight (OVE) began reporting annually on the performance of the Bank portfolio's evaluability in 2011.** The first report, which reviewed the evaluability of projects approved in 2011, built on Management's self-evaluation system as a way to assess the evaluability of the Bank's portfolio by validating the DEMs of a sample of projects. The first exercise was a pilot to set the standards OVE would use and to inform the Board and Management of the new approach. Since the DEM's criteria and process had just been revised in early 2011, OVE also used this pilot exercise to assess the DEM as an evaluability tool.
- 1.3 **The main findings of the first Evaluability Report were that OVE's scores were similar to, though slightly lower than, Management's scores for sovereign-guaranteed (SG) operations, while OVE was unable to validate the scores of non-sovereign-guaranteed (NSG) projects.** OVE recommended a full revision of the NSG DEM and some refinements in the SG DEM. Management has followed up on OVE's recommendations, implementing a few changes in the SG DEM in January 2013 and designing a new DEM for NSG projects that was recently approved by the Board.¹
- 1.4 **This report follows a similar approach, building on the self-evaluation system to assess evaluability by validating Management's DEM scores for a sample of SG projects approved by the Board in 2012.** OVE will wait until the new NSG DEM is implemented to include NSG projects in the evaluability review, thereby fulfilling the IDB-9 requirement to assess the evaluability of the Bank's overall portfolio. The main difference between last year's and this year's exercise is that less emphasis is given this year to the Risk Management section of the DEM, since one of the recommendations made by OVE and implemented by Management was the exclusion of this section from this determination of project evaluability.
- 1.5 **Two questions are addressed in this report: (i) How evaluable were 2012 SG projects according to the new DEM?, and (ii) Has the use of the SG DEM to**

¹ GN- 2489-8

date led to better project monitoring through the PMR during implementation? The DEM can be seen as a tool not only to assess project design but also to prepare projects for monitoring, ensuring they will have the information and institutional capacity needed for proper monitoring. In preparing this report, OVE reviewed about one-third of the portfolio approved in 2012 and validated their respective DEMs following the guidelines produced by Management in June 2012. In addition, OVE reviewed a sample of Progress Monitoring Reports (PMR) for projects approved before December 2011 and currently in execution.

II. EVALUABILITY OF SOVEREIGN GUARANTEED PROJECTS IN 2012

A. The Sample and Method of Analysis

- 2.1 The sample of projects reviewed in this report was selected at random among the SG projects approved in 2012, with each project having a one-third probability of being selected.** This method led to a sample of 41 projects, distributed across different sectors. As mentioned in the 2011 report, this sampling strategy is expected to generate a sample of projects representative of the Bank's portfolio as a whole, although it is not possible to make inferences by sector or by instrument for a given year.
- 2.2 As for 2011 projects, the evaluability scores in the DEM of each selected project were reviewed by an evaluability team (composed of a staff member and a research fellow) with knowledge of the relevant sector.** All teams were coordinated and overseen by a panel to ensure consistency. Teams used the SG DEM toolkit of June 2012 to rate project evaluability.
- 2.3 Management's evaluability scores for the sample analyzed by OVE are detailed in Figure 2.1 below and compared with the scores of the universe of projects approved in the same year.** The DEM's average score of 8.7 is the same for the sample and the universe. The maximum score in OVE's sample was 9.9, which also matches that of the universe, although the minimum is slightly higher (7.2 versus 6.6).² Figures 2.2 and 2.3 below show the comparison of the sample and universe by sector and instrument.

² Similar results were found in each individual section of the DEM, including Project Logic, Economic Analysis, and Monitoring and Evaluation.

Figure 2.1
Distribution of Management’s DEM Scores: Sample vs. Universe

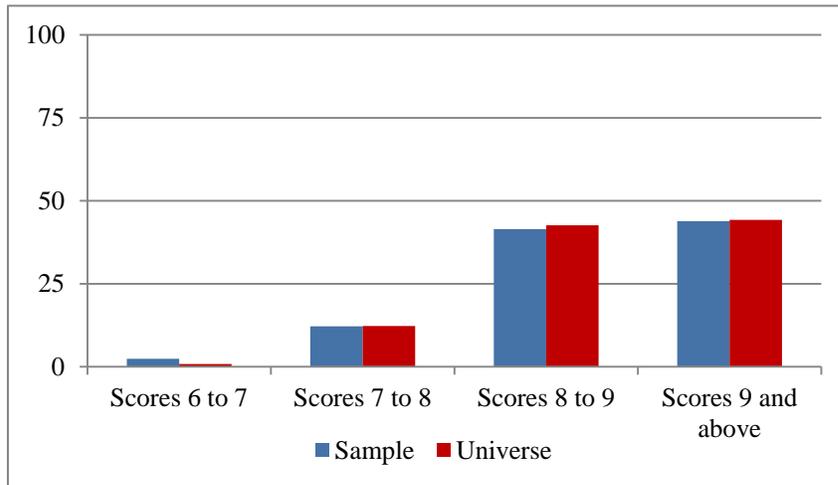


Figure 2.2
Projects by Instrument

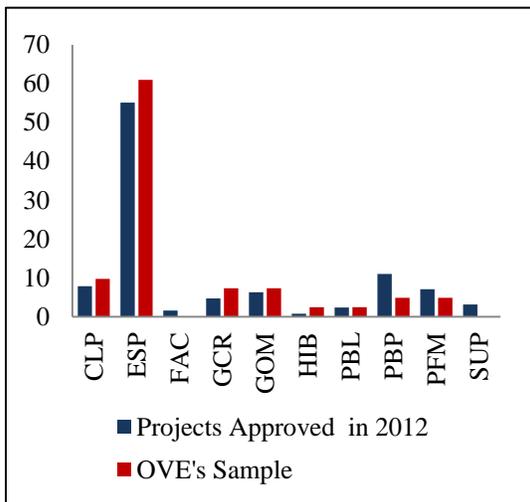
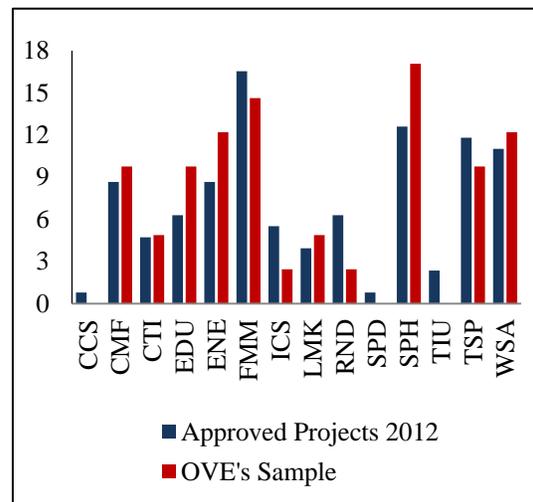


Figure 2.3
Projects by Sector



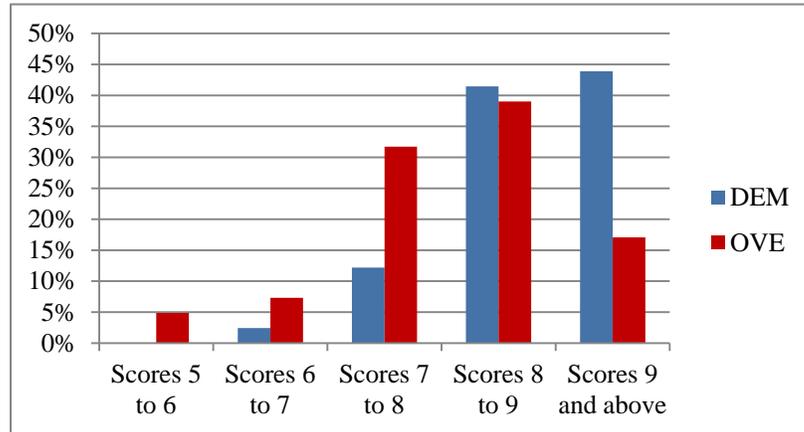
B. OVE’s Findings on Evaluability

2.4 **Figure 2.4 summarizes the differences in the distribution of overall evaluability scores between OVE and Management.** (See Box 2.1 for a description of criteria and Annex A for a list of all scores.) The average difference between Management’s and OVE’s overall evaluability rating was 0.7. Fewer than 25% of the projects had a difference bigger than one point, and in eight projects OVE’s score was higher than Management’s.³ Most of the differences were in the section on Project Logic, where OVE’s scores were lower in 80% of

³ These projects are EC-L1107, EC-L1113, HA-L1077, ME-L1103, ME-L1121, NI-L1059, NI-L1063, PE-L1092.

the projects. OVE's and Management's scores in other dimensions were generally similar.⁴

Figure 2.4
Distribution of Management's and OVE's Overall Evaluability Ratings
for 2012 Projects



⁴ In the M&E section the main disagreement was regarding project CH-L1064 (Program to Support SENCE's Effectiveness). As in the economic analysis, the impact evaluation plan presented by the M&E document did not propose an evaluation of IDB's intervention. Instead, it described the evaluation methodology that would be used to assess the two main programs carried out by SENCE (Programa de Capacitación en Oficios y Programa de Formación en el Puesto de Trabajo), which OVE did not consider acceptable. OVE's validation results agreed with Management that almost all projects fulfill all requirements in the Risk Matrix. As pointed out in the 2011 report, the Risk Management section is a very straightforward checklist, which verifies compliance of the Risk Matrix. There are four criteria in this section of the DEM: i) Identified risks have been rated for magnitude; ii) Identified risks have been rated for likelihood; iii) Major risks have identified proper mitigation measures; and iv) Mitigation measures have indicators for tracking their implementation. All projects in the sample included a risk analysis that classified their risks according to their probability and magnitude of the impact. Very few did not have a proper discussion of mitigation measures, and the main issue identified by OVE was the lack of indicators to monitor the mitigation measures (almost 15% of the sample).

Box 2.1
Evaluability in the SG DEM used in 2012

The second section of the DEM, which assesses evaluability, has four sections: (1) Project Logic; (2) Economic Analysis; (3) Monitoring and Evaluation and (4) Risk Management. Each section sums to ten and has equal weight in the final evaluability score.

Project Logic: This section has 33 checklist questions in three subsections: Program Diagnosis, Proposed Interventions or Solutions, and Results Matrix Quality. The purpose of this section is to analyze whether the Loan Document (LD) has a proper diagnosis, including identification of the development problems to be addressed, their root-causes, and how the intervention will address them. It also assesses whether the LD provides enough evidence of the effectiveness of the proposed intervention and a clear vertical logic (i.e. how inputs and activities will be translated into outputs, which will lead to outcomes, which will contribute to final impact). The Results Matrix sub-section asks about identification of SMART (Specific, Measurable, Attributable/Achievable, Relevant and Time-Bound) indicators, including baselines, targets and means of verification.

Economic Analysis: According to the criteria in this section, projects should present either a cost-benefit analysis (CBA) or a cost-effectiveness analysis (CEA). These should assess the economic benefits and costs of the intervention and potential alternatives. All assumptions should be explicitly mentioned and supported by evidence, and a sensitivity analysis should be included. If neither a CBA nor a CBE is included in the project, a proper justification should be presented.

Monitoring and Evaluation: All projects should comply with a minimum standard for monitoring and evaluation. The DEM includes 16 questions in this section regarding the necessary “*elements for projects to include systems to track implementation and measure results*” (GN-2489-4), such as definition of monitoring and evaluation plans, budget for the activities, and methods.

Risk Management: This section has four compliance questions related to the Risk Matrix. The first two questions ask whether the likelihood and magnitude of identified risks are measured. The last two questions relate to proposed mitigation steps and indicators for monitoring their implementation. All questions have the same weight.

Although all questions are binary, some require more judgment by the evaluator than others, such as: “*The main problem is clearly identified,*” or “*the main factors (or causes) contributing to the problem are clearly identified,*” or, following the tutorial for specific country characteristics required in question 6, “*(...) the POD includes, in a precise manner, the historical, political, social, and economic context in which the problem arose, as well as those in which its solution will be attempted.*”

Source: 2011 Evaluability Review of Bank Projects, RE-397-1

- 2.5 **In the few cases in which a large divergence was identified, the difference in scores was driven primarily by the Economic Analysis and Risk Management sections.** These two sections have only a few criteria each, and any disagreement can lead to high score differences. As discussed above, in two projects -- the Social Safety Net Support Program in Honduras (HO-L1071) and the Program to Support SENCE’s Effectiveness in Chile (CH-L1064) -- OVE judged the Economic Analysis to be inappropriate. Both projects support existing government programs, and their economic analyses focused on the programs themselves rather than on the value added of IDB’s support. In a few projects OVE disagreed with Management’s assessment of risk mitigation measures and/or related indicators. For instance, the risk matrix for CH-L1064 is

incomplete, identifying only an indicator related to fiduciary risk, which is classified as low probability and medium impact. Other risks, including those with higher probability and likely impact, lack indicators for risk mitigation.

1. Project Logic

- 2.6 **As in last year’s OVE report, most evaluability problems were related to weak program logic and lack of evidence to support the diagnosis presented in the loan document.** Indeed, OVE’s average scores on the Project Logic section (comprised of three subsections -- Project Diagnosis, Project Logic, and Results Matrix) were substantially lower, at 6.8, than the average scores for the other three sections (8.9, 7.5, and 9.4 for Economic Analysis, Monitoring and Evaluation, and Risk Management, respectively). This lack of evidence ultimately weakened both the overall logic of the project and the ability to define a proper strategy to measure results at completion.

Table 2.1
Percentage of Projects that did not fulfill DEM Criteria

The main problem being addressed by the project is clearly identified	15%
The intended beneficiary population is clearly identified (households, localities, firms, users, or overall population)	24%
The main factors (or causes) contributing to the problem are clearly identified	37%
Empirical evidence of the main determinants of the problem is provided	51%
Magnitudes of deficiencies are provided for main factors (in order to assess the relative importance of identified factors)	63%
Diagnosis takes into account specific country characteristics in the area of project intervention	17%

- 2.7 **OVE found that 15% of projects did not clearly identify the problem to be addressed (Table 2.1), and in many cases there were problems in the discussion regarding the main determinants of the problem.** Some of the main issues identified are the following:

- **Unclear distinction between the problem, its causes and its consequences.** For example, the Fiscal Consolidation Program for Guatemala (GU-L1064) presented a detailed description of the country’s fiscal management issues, but confused the problem being addressed (the country's fiscal deficit), the factors that caused it (low/unstable tax revenues associated with inadequate tax laws and high dependence on VAT; low institutional capacity of ministry of finance; poor fiscal

management at the municipal level), and the challenges resulting from it (low municipal financial capacity to provide public services).

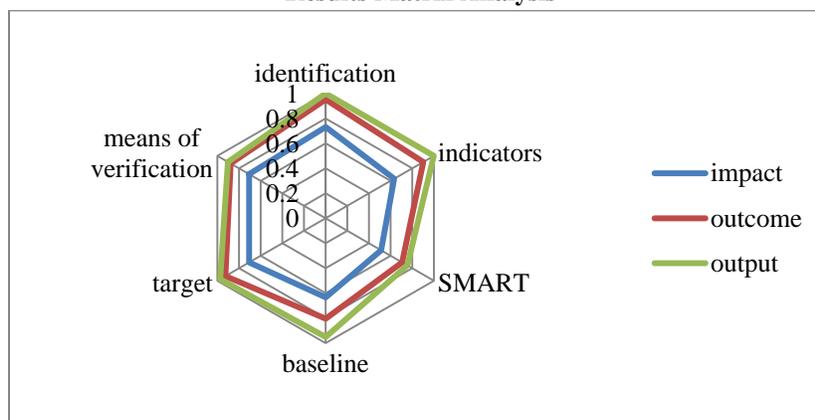
- **Incomplete diagnosis.** For instance, the Second Basic Education Improvement Program in Suriname (SU-L1019) – which focuses on developing curriculum for the basic education system and on increasing learning outcomes of students in grades 1 to 8 -- stated that current learning outcomes in the country were not satisfactory and presented some data on enrollment rates, public spending on education, and number of teachers, but did not provide evidence on the deficit in learning outcomes (e.g. scores on international or national tests). The loan proposal also mentioned differences in educational attainment between urban and rural/interior areas but did not provide information or evidence describing and quantifying these disparities.
- **Diagnosis based on outdated data.** That is, for example, the case of the rural Water Supply and Wastewater Management Program in Colombia (CO-L1105), which provided some information on the problem being addressed, such as the share of population lacking access to adequate sanitation in urban and rural areas. However, the data were somewhat outdated, mostly from 2008, and it was unclear whether changes had occurred since then that would affect the need for the project.
- **Lack of evidence on the main determinants of the problem.** For example, the Novo Hamburgo Integrated Municipal Development Program (BR-L1187) is a PROCIDADES program aimed at improving socioeconomic, urban and environmental conditions in the municipality of Novo Hamburgo. OVE and SPD agree that the problems being addressed in this program -- urban decay and deficient infrastructure in specific areas; low competitiveness of leather and footwear industry; and increase in violence -- were clearly identified in the loan proposal. However, OVE found that the main determinants of these problems were not sufficiently detailed, as the loan proposal listed general factors thought to contribute to the problems but did not provide evidence supporting those claims. Additionally, the document did not include information indicating the dimension or relative importance of these factors.

2.8 **The Project Logic score was also affected in many cases by lack of internal validity and vertical logic.** Proper evidence of the effectiveness of the proposed interventions was missing in almost half of the projects, and a slightly higher share of projects did not include convincing information about the applicability of the intervention. In many of those cases OVE also judged the logic to be unsuitably articulated, since the link between inputs, outputs, results and impact was unclear.

2.9 **The identification of outputs and outcomes in projects' results matrices – as well as their respective indicators – was better than that of impacts (Figure**

2.5), as was also true the prior year. All projects identified deliverables and indicators for them, though not all indicators were “SMART” or had means of verification. Almost all projects described the expected outcomes of the project, though a smaller share had proper indicators, including baselines and means of verification. As in last year’s report, teams had some difficulty adequately pinpointing impacts and finding appropriate indicators, and they sometimes confused outcomes and impacts. This might be due to the confusion created by the use of the term *impact*, as the term “impact evaluation” is generally used for assessing the outcomes of a project. As recommended by OVE last year, Management has agreed to remove the scoring of impact and associated indicators from the overall evaluability score from January 2013 on.

Figure 2.5
Results Matrix Analysis



Note: Each section of the radar graph shows the percentage of projects that scored positively in the respective criteria of the DEM for each of the three dimensions, impact, outcome and output, according to OVE’s analysis.

2. Economic Analysis

2.10 IDB-9 requires all projects submitted to the Board to have some sort of economic analysis. All projects approved in 2012 fulfilled this requirement. In the 2012 sample analyzed by OVE, two projects had a cost-effectiveness analysis,⁵ while all others had a cost-benefit analysis. The general quality of such analyses has improved in comparison to those in projects approved in previous years. Among the projects with a cost-benefit analysis, OVE judged that benefits and costs were properly identified in about 87% of the projects analyzed. OVE identified issues with the assumptions in 20% of projects, and only 12% (or five projects) failed to provide any sort of sensitivity analysis.

⁵ AR-L1127 included a cost-effectiveness analysis and PE-L1092 included both cost-effectiveness and cost-benefit analyses.

- 2.11 **Despite the general improvement in the economic analyses in projects, their levels of rigor were quite heterogeneous** As a positive example, the loan proposal for the Innovation and Human Capital for Competitiveness Program in Costa Rica (CR-L1043) included a comprehensive economic analysis for each component, with almost all including a sensitivity analysis. In contrast, the economic analysis for the Pilot Program for the Generation and Development of Productive Fabric in Priority Economic Development Areas in Haiti (HA-L1068) was very poor. The project aimed to catalyze the emergence of a sufficient supply of goods and services in the Northern Economic Growth Pole by supporting selected initiatives for agribusiness, tourism, and industrial value chains. The net present value of the project was based on unrealistic or unlikely assumptions, which was likely to lead to an overestimation of the project's economic returns. The analysis did not include a sensitivity analysis regarding the number of firms reached by the program, a key determinant of project effectiveness.
- 2.12 **Coverage of the economic analyses also varied, with some considering only a subset of project components and others considering the whole government program, including parts not supported by the Bank.** For example, the Comprehensive Support for Effective Labor and Social Security Policies in El Salvador (ES-L1063) included a detailed cost-benefit analysis; however it covered only one component of the project. The analysis in the Program to Support SENCE's Effectiveness in Chile (CH-L1064) focused on the costs and benefits of the activities currently carried out by SENCE, showing evidence that those programs provide positive net social benefit. However, the loan is intended to improve the administrative capacities of the institution, and there is no analysis on the expected benefits of this specific intervention. While the analysis presented may serve to justify the decision to help improve SENCE's effectiveness, it does not address the issue of whether IDB's intervention will bring about incremental benefits to the agency's programs that surpass the intervention's economic costs. A similar problem was identified in the Social Safety Net Support Program in Honduras (HO-L1071), where the cost-benefit analysis covered Bono 10,000 as a whole, including but not limited to the changes expected through this operation.

3. Monitoring and Evaluation

- 2.13 **Three-fifths of the 2012 sample proposed a rigorous impact evaluation – either through a random assignment or a non-experimental method -- compared to one-half in the prior year's sample.** Although this is a positive trend overall, OVE found the quality of proposals to be very heterogeneous. Some proposals did not discuss other interventions going on in parallel and potentially affecting the same beneficiaries, a situation which, if not properly addressed, could distort the results of the evaluation. An example is the program for Mother and Child Hospital Network Strengthening in Honduras (HO-L1072), where the lack of information on how other initiatives on maternal health in the country complement and/or overlap with the current project might invalidate the identification strategy proposed in the impact evaluation. In addition, the program includes many dissimilar components, and the evaluation proposal did not indicate how to assess

the relative impact of each component on the overall result, though a null (or even undesirable) result of one component could potentially result in the underestimation of the contribution of another component.

C. Classification of Project Evaluability

2.14 **The IDB-9 Agreement requires an evaluability score of at least 5 as a threshold for project approval.**⁶ SPD's classification system, used to track progress over time in the DEO, labels each project as Highly Evaluable (score ranging from 7 to 10), Evaluable (score between 6 and 7), Satisfactory (score from 5 to 6) or Unsatisfactory (score below 5).⁷

2.15 **In 2012, 99% of IDB's projects were classified as Highly Evaluable, despite heterogeneity in their ex ante ability to demonstrate results at completion.** This indicates that the current classification is not particularly informative nor a strong tool in tracking progress over time. It may also be misleading, as in a few cases the evaluability notes prepared by SPD indicated problems that were not reflected in the final scores and that seemed incompatible with the "Highly Evaluable" label given to the project. For instance, in the project Support for Productive Community Secondary Education in Bolivia (BO-L1071), SPD's note identified a number of problems that seriously affect evaluability: "*The diagnostic identifies several factors that contribute to the problem and empirical evidence based on the population of interest is provided. The intervention is clearly linked to the problem and evidence is provided in other context. No evidence on the applicability to Bolivia is provided. The intended beneficiary population is not clearly defined. The students and schools that will benefit depend on the availability of municipalities to commit resources for the operation. This item is categorized as one with high risk in the risk matrix and therefore the identification of beneficiaries is not considered clearly defined.*" However, since the overall score was 8.4, this project was still classified as Highly Evaluable.

2.16 **Projects with evaluability scores between 7 and 8 are most likely to have deficiencies in their ex-ante ability to report on results that are incompatible with the classification of "highly evaluable".** SPD's evaluability note for the Multisector Preinvestment Program IV (AR-L1149), which is the second tranche of a CCLIP aimed at helping increase public investment by national and subnational governments in Argentina, stressed that "[...] *the analysis carried out*

⁶ According to the "Report on the Ninth General Increase in the Resources of the Inter-American Development Bank" (AB-2764) "*In cases of extraordinary humanitarian or financial crisis (acute payment pressures), and upon the presentation of a properly justified request by the President, the Board of Executive Directors may waive the required threshold before approving an operation, as long as the request is accompanied by a timetable giving the shortest possible time period, not to exceed 90 days, for achieving the said threshold.*"

⁷ Since 2011 no SG project with score below 5 has been considered by the Board, as projects with evaluability score below 5 are required by Management to be improved prior to Board consideration. These projects in the past were classified as Unsatisfactory.

states that economic growth requires additional public investment, particularly at sub-national level. This intervention logic is associated with the vertical logic of the project: pre-investment studies → increased and better quality executable portfolio → incremental investments (at national, provincial and municipal). However, no evidence is presented regarding the impact of the first operation of this CCLIP. Not all indicators at outcome levels are SMART.” Additionally, the note points out problems with the economic analysis: “[t]he project has a cost-benefit analysis where costs and benefits are not all properly identified and quantified.” These issues were, to a certain extent, reflected in the respective sections’ scores – 6.0 in Section 3 and 5.5 in Section 4. Nonetheless, the high score in Section 5 inflated the project’s overall evaluability score, which was 7.2, leading the project to be classified as Highly Evaluatable despite the significant evaluability issues it presents.

- 2.17 **Another issue with this classification, as mentioned in the 2011 report, is the inflation of ratings caused by the inclusion of the scores from the Risk Management section.** From January 2013 on, the score on Risk Management will not affect the overall evaluability score of projects, but for the 2012 sample analyzed here it had a significant impact. The elimination of Risk Management from the 2012 scores would have resulted in a downward shift of about 10 percent in the number of projects rated as Highly Evaluatable.

III. PROJECT MONITORING AFTER THE DEM

- 3.1 **In 2009 the Bank launched the second tool in the SG Development Effectiveness Framework - DEF, the Progress Monitoring Report (PMR), to be used for all projects to report on implementation progress. OVE sought to assess the extent to which the introduction of the DEM helped projects prepare for monitoring during implementation.**⁸ The PMR has a very simple premise: once a project is designed with a Results Matrix -- including indicators with metrics, baselines, milestones, and targets -- and their respective costs have been estimated, it should be easy to monitor progress vis-à-vis the initial plan, provided that no changes are made during execution.
- 3.2 **Changes in project outputs and outcomes can break the link between the assessment in the DEM and the PMR.** In some cases changes do need to be made, and these changes in turn lead to adjustments in the monitoring and evaluation plan. In addition, the New Project Cycle allows some of the analysis to

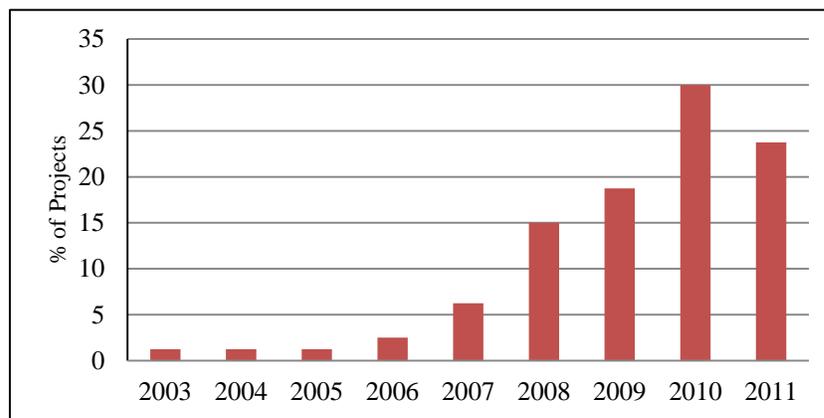
⁸ The analysis took into account all the instruments used in the Bank to prepare and implement projects, including the Results Matrix, the Monitoring and Evaluation Plan, and the Annual Operation Plan (AOP). The Results Matrix is the instrument used by the Bank to summarize the vertical logic of the project, defining the outputs, intermediate outcomes, final outcomes and impacts of the project. The Project Execution Plan, not required as an annex, establishes the relation between the products of the project and the allocation of resources by component and subcomponent, and is the basis for the AOP and the Acquisition Plans.

be done in the first months of execution, which can lead to refinement of indicators after approval. In these cases it is important to track any alterations to the Results Matrix included in the approved Loan Proposal, in order to maintain accountability.

A. Sample and Method of Analysis

3.3 **OVE randomly selected a sample of 80 projects that had less than 10% disbursed as of December 2009 or had been approved since then, (about 20% of the universe) and reviewed the March 2011 and March 2012 PMR cycles.** The sample was divided into two groups: (a) projects approved before April 2009 (which did not have a DEM), and (b) projects approved between April 2009 and December 2011 (all of which had a DEM at approval). Figure 3.1 shows the distribution of projects by year. OVE recognizes that the analysis is only broadly indicative given the many other changes in the Bank in recent years.

**Figure 3.1
Distribution of Projects in the sample by year**



B. Main Findings

3.4 **Table 3.1 shows the results from OVE’s assessment of the PMRs approved before and after April 2009, when the DEM was introduced.** Though there were no major differences between the PMRs of projects approved before and after the DEM (possibly due in part to SPD’s work to help project teams improve their PMRs and correct problems from design⁹), PMRs for projects approved after the implementation of the DEM had slightly better output monitoring, including the definition of the indicator and data on actual values and their respective costs.¹⁰

⁹ The background evaluation of the DEF and DEO prepared by OVE for the evaluation of the IDB-9 agreement, Background Paper: The Development Effectiveness Framework and the Development Effectiveness Overview, shows evidence that PMRs have improved over time.

¹⁰ The differences for all the categories in table 3.1 are statistically significant at level of significance of 5%.

Table 3.1
Assessment of Outcome and Outputs in the PMR

Criteria	Projects Approved Before 4-09	Projects Approved After 4-09
The outcome is a genuine outcome - neither an impact nor an output	85.6%	82.5%
Considering all reported outcomes, the outcome comes from the Results Matrix, or the changes are explained in the issues section	50.6%	60.6%
Considering only genuine outcomes, the outcome comes from the Results Matrix	99.9%	76.4%
The outcome indicator contributes to measure the outcome stated	93.1%	91.2%
Each output represents a single good or service	86.1%	90.4%
An actual value is reported for the output, when expected*	96.6%	99.4%
The output indicator is measurable	65.3%	68.0%

Note: Since projects have different number of outcome and output indicators, the assessment was done by indicator and an average result per project was computed.

* This statistics were generated only considering the indicators that are measurable.

3.5 The definition of outcome appears to have worsened slightly after the implementation of the DEM. This could have been influenced by the requirement that teams identify impacts as well as outcomes, leading teams to shift the classification of outcomes to impacts and consequently define some outputs as outcomes. While almost all outcomes included in the Results Matrix of projects approved before 2009 were judged to indeed be outcomes, the same is true for only about 76% the projects approved after 2009.

3.6 Although higher evaluability scores might be expected to lead to better PMRs, this was not generally the case in OVE's sample. It is true that higher DEM scores for indicators were correlated with more measurable indicators in the PMR, but in most cases the indicators in the LD and in the PMR were not the same.¹¹ The Multiphase Rural Electrification Program in Guatemala (GU-L1018)¹², for example, scored high on the quality of indicators in both the DEM and the PMR. The desk review revealed, however, that during project implementation several problems of design were detected and addressed, and the

¹¹ A large share of projects has some changes in the indicators from the Results Matrix to the PMR. In many cases the change was related to an improvement in the indicators, but without changing their essence. In other cases, the changes in the project led to the definition of completely different indicators.

¹² This project was approved in 2008 prior to the implementation of the DEM. So the team who designed the project did not have the pressure of the standards set in the DEM. During the preparation of the 2008 – 2009 DEO, SPD prepared DEMs for all projects approved in 2008 based on their documents at approval.

indicators were changed. On the positive side, the monitoring of the problems faced by the project during its execution allowed the team to design a plan to overcome them.

3.7 **In-depth analyses of five projects¹³ suggest that a number of challenges faced by project teams might explain the mixed results (see Box 3.1 for examples):**

- **Some projects must be modified to adapt to changes in the countries.** Most projects experienced major changes once implementation began. Unexpected modifications to regulations, changes in the government, or cuts in the budget forced rethinking of the project. Unfortunately, teams rarely document these unexpected changes and their impact on project execution. Teams have space in the Issue section of the PMR to report on changes to the Results Matrix, but OVE found that this did not occur in almost two-thirds of the relevant cases.
- **A shortage of time during the design phase can lead teams to focus on other parts of the DEM, giving less attention to the section on the PMR.** More than 45% of the loan documents for projects in the 2012 sample did not specify costs by outputs, and only 18 out of 122 appeared to have a draft PMR at approval.¹⁴ These projects had generally higher DEM scores, particularly on the criteria related to the Results Matrix and PMR. Although it is not a formal requirement that teams should prepare a draft PMR before the start of project implementation, it does help teams ensure that projects have the necessary elements for monitoring through the PMR.
- **Different staff members may be responsible for different aspects of project preparation (POD and Results Matrix, Annual Operation Plan -AOP, Project Execution Plan - PEP), without sufficient oversight to ensure consistency.** As mentioned above, the basic elements that will later feed into project monitoring are prepared at design. OVE noticed inconsistencies in some projects between the content and structure of the Loan Proposal, AOP, PEP, Evaluation and Monitoring Plan, and Procurement Plan. In some cases these inconsistencies were transferred directly into the PMR.
- **Project teams receive limited guidance on how to prepare project documents that lay out the framework for subsequent monitoring.** OVE's interviews indicate that project teams receive limited guidance and

¹³ Out of the 80 projects in the sample, five were analyzed in depth and all their documentation was reviewed, including the POD, its annexes and the PMRs. These five projects were chosen at random to exemplify cases in which the DEM but not the PMR had high scores, or vice-versa, and one case with the DEM and PMR scoring similarly. The projects selected were BR-L1020, BR-L1026, EC-L1075, ES-L1045, and GU-L1018.

¹⁴ Those 18 projects included a draft PMR among the documents in the annex.

training on how to prepare the annexes in the POD, and the outline for the preparation of the POD has limited and outdated information. For instance, at the end of the outline in VPS' webpage is some information on how to prepare a Results Matrix. Important elements for the PMR are not contemplated in this outline, such as the need to identify means of verification and units of measure for the indicators. The general information provided is vague and has not been revised since the implementation of the PMR.

- **Teams often change during execution.** Though team leaders are supposed to remain responsible for the implementation of projects after approval, in practice the composition of teams often changes over time. Though new team leaders are supposed to follow the monitoring and evaluation plan approved with the project, OVE's interviews suggest that in many instances new team leaders make unexplained changes in PMR indicators that are not consistent with the approved Results Matrix.

Box 3.1
Changes in Project Monitoring Plans from Approval to Execution

OVE analyzed a subset of projects in greater detail and found many cases where information in PMRs varied from that contained in the loan documents approved by the Board. In some cases these changes incorporated new information, strengthened indicators, or filled gaps left at design, leading to better project monitoring. In other cases, the changes resulted from changes in team leaders, in the government, or in project context, and had a negative impact on project monitoring. Two examples of each are outlined below.

The main objective of the National Infrastructure Program for the Universalization of Education with Quality and Equity (EC-L1075) in Ecuador is to increase education coverage in *parroquias* with high rates of unmet basic needs and limited educational opportunities. Although the indicators were deemed SMART in the DEM diagnostic, in the PMR many of them were identified as not being measurable. The main challenge faced by the team, whose leadership changed during execution, was the many adjustments experienced by the project. The project was approved in 2010, but for many reasons the executing agency delayed the project, which is currently being considered for reformulation. This led to the need to redefine the indicators and the monitoring strategy.

The program for Local Cluster Competitiveness in the state of Pernambuco, Brazil (BR-L1020) also received high scores in the DEM and low scores in the PMR. The desk review of this project pointed out apparent inconsistencies in structure among the LD, the Procurement Plan, the Plan of Activities, and the Logical Framework.

In contrast, the Program for Rural Roads for Development in El Salvador (ES-L1045) and the Integrated Social and Urban Development Program of the Municipio of Paranagua (BR-L1226) received higher scores in the PMR than in the DEM. In the case of ES-L1045, the process of preparing the PMR filled gaps in the project at design by including better and more detailed outcomes and linking the roads in the projects to their corresponding planned values and costs. According to interviews with the team, SPD has had an important role in helping the team improve the quality of reporting. Nevertheless, the PMR did not justify the changes in the target values for the kilometers to be improved or rehabilitated by the project, or the delay in the production of the first outcomes by one year. Both this project and EC-L1075 had a change of team leader during execution.

BR-L1226 is an example of how problems at design can translate into problems at the monitoring stage. The desk review of the project confirmed problems in (i) the definition of the main impacts, outcomes and outputs, (ii) the scope of the monitoring and evaluation plan, and (iii) inconsistencies and lack of specificities in the acquisition plan. These problems undermined the capacity of the loan documents to serve as inputs for proper monitoring through the PMR. Although the PMRs enhanced the structure and information in the Results Matrix, this effort has not materialized into better monitoring of the project. Continuous changes in the information presented in the PMR, the absence of planned and actual values of the outputs and costs, and changes in regulations affecting project targets have weakened the PMR as an effective monitoring tool. On a positive note, the comments in the Issues section of the PMR have helped explain project dynamics.

- 3.8 **Many staff interviewed by OVE recognized the value of SPD's work in helping teams improve their PMRs, and of KNL's training, *Project Management for Results (PM4R)*.** The PM4R training considers all instruments used in the Bank relevant to project monitoring, from the Execution Plan to the Results Matrix to the Risk Matrix. The impact of this training will be more

evident in the future, as it started in 2011 and relevant projects are still in execution.

IV. CONCLUSIONS AND RECOMMENDATIONS

- 4.1 **This report reviews the evaluability of a sample of SG projects approved in 2012 by validating Management’s DEM scores.** As with the 2011 sample validated by OVE last year, OVE’s validations led to some lowering of Management’s scores, particularly in the section related to project logic, which typically was the weakest of the four sections of the DEM that compose the overall evaluability score. Despite the problems of evaluability identified in many projects, almost all projects continued to be classified, according to Management’s classification system, as “highly evaluable”, indicating that the way projects are currently labeled is not very informative and needs to be revised.
- 4.2 **The report also analyzes whether the introduction of the DEM in 2009 -- and the strength of DEMs since then -- translates into stronger PMRs for project monitoring during implementation.** It finds that projects approved after the implementation of the DEM have slightly better tracking of outputs (though not outcomes), but it fails to find a significant connection between better DEMs at entry and better PMRs during implementation. For a number of reasons -- whether because of problems in the original plan or because of changes in the project itself -- project teams often change project monitoring plans after project approval, diverging from the original project monitoring and evaluation plan either positively or negatively. On the positive side, some PMRs have been improved through SPD support even given imperfect monitoring plans at design.
- 4.3 **OVE offers the following five recommendations for Management:**
1. **Focus greater attention on ensuring proper problem diagnosis at the project design stage.** As in the 2011 review, incomplete problem diagnosis and inadequate project logic are the main weaknesses in the evaluability of SG projects at entry, and these can make it very difficult to be able to measure actual achievements at completion.
 2. **Revise the classification system for projects’ evaluability** (as reported, among other places, in the Development Effectiveness Overview - DEO). The current classification is non-informative, since almost all projects fall in the same (and somewhat misleading) category of “highly evaluable”. A more elaborated classification should be considered, which applies narrower ranges and more accurate labels for each category.
 3. **Strengthen SPD’s evaluability notes to better explain the DEM scores.** The evaluability notes should provide more information regarding the evaluability issues identified in the project and be given more weight in the overall assessment of evaluability.

- 4. Increase integration between the documents produced during project design (POD, AOP, PEP, and Results Matrix) and the PMR, and provide adequate guidance and training to ensure their proper preparation by project teams.** Teams need clear guidance in preparing project documents at both the design and monitoring stages in order to improve their overall quality and minimize the need for changes over time.
- 5. Ensure that the information from the approved Results Matrix is entered in the system prior to implementation and any changes during implementation are tracked.** Preparation of PMRs at the design stage can help teams achieve better DEMs and make changes in indicators after approval more visible. Ensuring that any changes and the need for them are well documented from the beginning can enhance accountability and provide information for future projects.

ANNEX A. PROJECTS SELECTED FOR VALIDATION

Project Number	Name	Type	Sector	Eval Score
AR-L1068	Program for Rural Development and Family Agriculture, PRODAF	ESP	RND	9.8
AR-L1127	Program for Strengthen. Provincial. Inst & Fiscal Mng. Stage II (PROFIP II)	ESP	ICS	7.9
AR-L1136	Development Programme's Norte Grande provinces: W&S Infrastructure	GOM	WSA	8.5
AR-L1149	Multisectorial Preinvestment Program IV (CCLIP)	CLP	FMM	7.2
BO-L1064	Grow Well to Live Well Early Childhood Development Program	ESP	SPH	8.9
BO-L1071	Productive Communitarian Secondary Education	ESP	EDU	8.4
BR-L1187	PROCIDADES - Novo Hamburgo	GCR	FMM	8.7
BR-L1190	PROCIDADES – Recife	GCR	FMM	8.1
BR-L1272	Blumenau's Sustainable Urban Mobility Program (BID-Blumenau Program)	ESP	TSP	9.3
BR-L1287	Social Inclusion and Opportunities for Youth in Rio de Janeiro	ESP	SPH	8.5
BR-L1299	Program to Strengthen the Social Assistance System	ESP	SPH	8.8
BR-L1303	CEEE Generation and Transmission Project	ESP	ENE	9.3
BR-L1314	Mauês Integrated Sanitation Program - PROSAIMAUÉS	ESP	WSA	8.5
BR-L1336	Santa Catarina Logistics Infrastructure Program	GOM	TSP	9.1
BR-L1337	Fiscal stability consolidation program for the development of the State of Bahia	PBL	FMM	8.5
CH-L1064	Program to Support SENCE's Effectiveness	ESP	LMK	9.8
CO-L1093	Support Program to Quality Education Plan for Prosperity	ESP	EDU	8.7
CO-L1105	Rural Water Supply and Wastewater Management Program	ESP	WSA	8.4
CR-L1043	Innovation and Human Capital for Competitiveness Program	ESP	CTI	9.5
CR-L1049	Power Sector Dev. Program 2012-2016 (Reventazon Hydroelectric Project)	CLP	ENE	9.0
EC-L1107	Program to Support Integrated Early Childhood Development	ESP	SPH	9.1
EC-L1111	Quito Metropolitan Urban Transport System	ESP	TSP	9.3
EC-L1113	Social Housing National Program - Stage 2	ESP	FMM	7.6
ES-L1063	Overall support to the effectiveness of labor and social security policies	ESP	LMK	9.5
GU-L1064	Fiscal Consolidation Program for Guatemala	HIB	FMM	7.6
HA-L1068	Northern Economic Pole Business Accelerator Program	ESP	CMF	7.9
HA-L1073	Institutional Transformation and Modernization Program of the Energy Sector II	PBP	ENE	9.1
HA-L1077	Increasing Access to Quality Education in Haiti	ESP	EDU	9.3
HO-L1071	Support to the Social Protection Network Program	ESP	SPH	9.1
HO-L1072	Mother and Child Hospital Network Strengthening Program	ESP	SPH	9.3
ME-L1103	Global Credit Program for Development of Mortgage Markets IV	CLP	CMF	7.4
ME-L1120	Financing Low Carbon Strategies in Forest Landscapes	GCR	CMF	8.5
ME-L1121	CTF-IADB "ECOCASA" Program	CLP	CMF	7.7
NI-L1052	Support to the Transportation Sector II Program	GOM	TSP	8.7
NI-L1059	Program to accompany the implementation of the National Policy for children in	ESP	SPH	9.1
NI-L1063	National Sustainable Electrification and Renewable Energy Program III	ESP	ENE	8.9
PE-L1068	Innovation Project for Competitiveness	ESP	CTI	9.3
PE-L1092	Project for the Dev. of Solid Waste Management Systems in Priority Areas	ESP	WSA	8.4
PE-L1121	Development of a New Sustainable Energy Matrix Program IV	PBP	ENE	9
SU-L1019	Second Basic Education Improvement Program (2nd BEIP)	PFM	EDU	9.9
TT-L1026	Multi-Phase Wastewater Rehabilitation Program- Phase I	PFM	WSA	9.1

Note: About one-third of the 2012 SG projects were selected at random to be validated, as each project approved had a one- third chance of being selected.