

The Quality of Results Frameworks in Development Policy Operations

IEG Learning product

June 24, 2015



Contents

ABBREVIATIONS	III
ACKNOWLEDGMENTS	IV
OVERVIEW.....	V
1. INTRODUCTION.....	1
2. METHODOLOGY.....	3
3. PRESENTATION OF RESULTS FRAMEWORKS	5
Overview of Recent Changes in OPCS Guidelines	5
Issues with the Results Framework Presentation	6
Presentation in Legal Documents Underpinning DPOs	10
4. QUALITY OF PRIOR ACTIONS	11
Relevance	12
Criticality.....	12
Additionality	15
Measurability	16
Efficacy	16
Other Considerations in Quality of Prior Actions	18
Summary on Quality of Prior Actions	19
5. RESULTS FRAMEWORKS IN PROGRAMMATIC DPOS	20
6. LESSONS FOR DPO MONITORING FRAMEWORKS	23
7. QUALITY OF REPORTING AT COMPLETION AND IMPLICATIONS FOR VALIDATION AND EVALUATION.....	25
Lessons from ICRs.....	25
Implications for IEGs Validation and Evaluation Methodology	26
REFERENCES.....	28
APPENDIX A. RESULTS IN DEVELOPMENT POLICY OPERATIONS.....	29
APPENDIX B. EXAMPLE OF POLICY MATRIX.....	31

CONTENTS

Boxes

Box 3.1. Examples of Lack of Explicit Statements of Outcomes that Undermine the Clarity of the Results Framework.....	8
Box 3.2. A Model of Best Practice Results Framework Presentation in the Policy Matrix of Development Policy Operations	9
Box 4.1. Operations Policy and Country Services' Classification of Prior Actions	13
Box 4.2. Use of Prior Actions on Laws.....	14
Box 5.1. An Example of a Partially Met Trigger without Institutional Depth	21
Box 5.2. Example of Weakening of Design in Programmatic Series.....	22

Tables

Table 2-1. Coverage and Methodology Used in Evaluative Questions	4
Table 4-1. IEG Criteria for Assessing the Prior Actions of Development Policy Operations	11
Table 4-2. Status of Draft Laws Not Submitted to the Parliament used as PAs.....	18

Figures

Figure 1-1. Outcome Rating and Relevance of Design Ratings for DPOs.....	2
Figure 3-1. Presentation of Results Framework in DPOs Policy Documents.....	7

<i>Evaluation Managers</i>	
❖ Caroline Heider	Director-General, Evaluation
❖ Nicholas David York	Acting Director, Independent Evaluation Group, Public Sector
❖ Mark Sundberg	Manager, Independent Evaluation Group, Public Sector
❖ Aghassi Mkrtchyan	Task Manager

Abbreviations

CAS	Country Assistance Strategy
CPIA	Country Policy and Institutional Assessment
DPO	Development Policy Operation
HIV/AIDS	Human Immunodeficiency Virus/Acquired Immune Deficiency Syndrome
ICR	Implementation Completion and Results Report
ICRR	Implementation Completion and Results Report Review
IEG	Independent Evaluation Group
M&E	Monitoring and Evaluation
OPCS	Operations Policy and Country Services
PA	Prior Action
PPAR	Project Performance Assessment Report
PRSC	Poverty Reduction Support Credit
RF	Results Framework

Acknowledgments

This report by the Independent Evaluation Group (IEG) was led by Aghassi Mkrtchyan (task manager) under the supervision of Mark Sundberg (Manager, IEGPS2) and the general direction of Nicholas York (Acting Director, IEGPS) and Caroline Heider (Director General, IEG).

The analysis was conducted by the IEG team consisting of Malathi Jayawickrama, Aghassi Mkrtchyan, and Jesse Torrence. The task team is grateful to Mark Sundberg for substantive guidance and advice in all stages of the study and to Iradj Alikhani (Consultant), Zeljko Bogetic (Lead Economist), Kenneth Chomitz (Senior Adviser), Souleymane Coulibaly (Program Leader), Stephen Hutton (Evaluation Officer), Anjali Kumar (Lead Economist), Lourdes Pagaran (ICRR Coordinator), Marcelo Selowsky (Consultant), and Clay Wescott (Consultant) for their advice and comments on the draft report.

The efforts of team members who prepared case studies are appreciated: Javier Bronfmann, Malathi Jayawickrama, Moritz Piatti, and Marcelo Selowsky. Team assistance by Aline Dukuze and Yezena Zemene Yimer is gratefully acknowledged.

Overview

This review on the Results Frameworks (RFs) of development policy operations (DPOs) is part of a series of assessments to gain new insight into the factors that influence the design, implementation, and performance of policy lending of the World Bank. The review aims to help fill gaps in the Bank's knowledge about the results orientation of DPOs and implications for achieving institutional change that improves a borrower's sustained growth and poverty reduction outlook. It synthesizes and reexamines existing evidence on the quality of RFs and its contribution to the success of lending programs. It also derives lessons from the recent use of DPOs by focusing on the identification of successful operational arrangements that may improve a program's overall outcomes. In addition, it aims to strengthen learning in the Bank and the Independent Evaluation Group (IEG) and to influence debate on the design and use of DPOs, especially as the Bank prepares the next DPO retrospective.

Through the following questions, the review addresses the key evaluative query about the overall quality of RFs for DPOs:

- Do RFs present a clear statement of program objectives and outcomes? Do prior actions have sufficient institutional depth and criticality to achieve the targeted results?
- What are the implications of the programmatic approach for RFs?
- Does the monitoring and evaluation (M&E) framework capture the impact of DPO on country outcomes?
- Do ICRs provide a comprehensive account of attribution of the Bank's inputs to programs results?

This review also documents recent changes in the Bank's approach to RFs for DPOs, examines their impact on the quality of the instrument, and seeks to derive lessons for IEG on improving validation and evaluation of DPOs.

The main conclusions include:

- **The presentation of RFs for DPOs has changed substantially in recent years with mixed results.** Some of these changes, such as greater clarity of reform programs in the programmatic series, streamlining of M&E, improved statement of objectives, and greater time

allocated for reporting on results, address shortcomings affecting DPOs. Other changes aimed at streamlining, such as elimination of "benchmarks" and intermediate outcome indicators, might have improved the consistency in the presentation of RFs. But these changes have also substantially increased the role of the remaining RF elements—prior actions and outcome indicators. In some cases streamlining seems to have created confusion with respect to the clarity of statements of outcomes. The review notes that for the Bank's streamlined and simplified RF model to be effective, it is critical to ensure (i) clarity of presentation of objectives and outcomes, (ii) high quality of prior actions, and (iii) comprehensive reporting of results and their attribution to the Bank's inputs at completion.

- **Some DPOs suffer from lack of clear statements of objectives and outcomes.** This shortcoming is present in some newly approved DPOs that follow the modified template of Operations Policy and Country Services (OPCS). Their RFs lack explicitly stated outcomes, while result indicators in many cases fall short of meaningfully measuring a DPO's impact. Some straightforward solutions are available to improve the RFs' presentation, and DPO results orientation and evaluability. The report suggests ways

to address weaknesses in the presentation of the “theory of change” in DPOs.

- **The quality of prior actions is critical for the robustness of the RFs of DPOs.** The review highlights recent improvements in prior actions and suggests further steps to improve their quality:
- Avoid actions that do not support significant policy changes such as draft regulations at early stages of preparation (before approval by governments), agency level actions with little or no tangible implication for overall policy, statements of intentions, repeated prior actions on recurrent government functions that lack additionality, and “pilot” actions without a clearly defined scaling-up strategy.
- Avoid policy actions unrelated to Bank engagement with the client country, which is contrary to the Bank’s approach to budget support and undermines the additionality of DPOs. The report suggests formulating guidance and standards on prior actions to improve the results orientation of DPOs.
- Excessive use of flexibility in a programmatic series can compromise a DPO’s focus on results. Dropping essential triggers or accepting partially met triggers that do not capture the true essence of intended reforms may substantially undermine the quality of RFs. Although maintaining flexibility in DPOs is important, in many cases a better balance between flexibility and rigor would improve a DPO’s focus on results. The review also notes that recent changes in the presentation of medium term reform programs in a programmatic series have improved the clarity of RFs.
- **Recent changes in M&E frameworks have contributed positively to the overall clarity of RFs.** These changes include streamlining of the M&E framework and better reporting of actual values at completion. The greatest remaining challenge is the use of output oriented results indicators that affect
- M&E quality. Such indicators also affect the clarity of RFs in cases where there is

no explicit statement of outcomes in the policy matrix.

- The Implementation and Completion Results Report (ICRs) of DPOs need to focus more on the quality of prior actions and provide a more comprehensive account of policy changes triggered by them. To enhance the quality of reporting on results, ICRs should reconstruct and document the implicit results chain leading from prior actions to program results. The review finds that the extension of time allocated to ICR preparation from six to 12 months may enhance the quality of reporting and a DPO’s evaluability. It also effectively expands a DPO’s time horizon. To ensure sustainability of reform, the long-term monitoring of reform areas supported by DPOs can be integrated into the monitoring systems of both country assistance strategies and country partnership strategies.
- **IEG’s validation approach captures most issues related to RFs, but some modifications are needed.** First, IEG’s validation needs to be aligned better with the evolution of the instrument, especially in areas such as the Bank’s approach on the use of benchmarks and intermediate indicators. Second, IEG’s validation approach should be more prior-action focused in its assessment of relevance of design to reflect the recent changes in the instrument’s RF. The method developed for this study that examines the overall quality of prior actions through assessing their relevance, criticality, additionality, and monitorability can be used for streamlining IEG’s assessment of the relevance of design and the M&E framework. To address the key issue of attribution in IEG reviews, a stronger focus needs to be placed on assessing the links between prior actions and program’ results.
- **Addressing shortcomings in the RFs of DPOs’ are important to improve the results orientation of DPOs.** Some of these shortcomings related to the clarity and statements of RFs and each of its

components will be easy to address. Others, such as the rigor of reform programs of borrowers and resulting quality of prior actions, would require sustained focus at the corporate level.

OPCS and IEG coordination is important to improve and further harmonize reporting and validation standards of DPOs.

1. Introduction

1.1 The objective of this review is to synthesize existing evaluative evidence and generate new evidence on the quality of the Results Frameworks (RFs)¹ of development policy operations (DPOs). The study is organized around a number of evaluative questions on various aspects of DPO design, implementation, reporting, and evaluation. Its focus is on the quality of RF, taking objectives and coverage for granted. The study does not cover the scope and coverage of reform programs underpinning DPOs, including the relevance of objectives, or the choice of policy areas and the links between DPOs and country assistance strategies. It generally does not look at program achievements and outcomes, except in cases when it is required for assessing the quality of prior actions (PAs).

1.2 The RF articulates how the reforms and policy actions supported by the World Bank will contribute to and influence country development outcomes². In DPOs, the RF is presented in the policy matrix and should include a clear statement of the objectives, outcomes, and outcomes indicators to be achieved through policy and institutional changes triggered by PAs. As part of the overall drive for results, the Bank has strengthened its focus on the quality of RFs because it considers them critical to the outcomes of DPOs.

1.3 The design of DPOs' RFs and their evaluation remain challenging³. It is difficult to assess both the extent of policy and institutional change as well as the attribution of outcomes to the Bank's intervention through DPOs. The main metric used for assessing DPOs' effectiveness in promoting desirable country outcomes is the outcome rating assigned to the series of DPOs or stand-alone operations series at completion, which is validated by the Independent Evaluation Group (IEG). DPOs do reasonably well on that account (with around three-quarters of DPOs closed in FY10–13 assessed positively), but there is mixed evidence on the overall performance of the Bank's clients with respect to policy and institutional change for poverty reduction and growth. Evidence remains mixed on the key formula of the Bank's' DPOs – external financing for policy and institutional change (see appendix A).

1.4 The RFs for DPOs is subject to various corporate level assessments. In the 2012 DPO Retrospective, OPCS assessed the quality of RFs through assessment of PAs and

¹ This learning product belongs to a series of three learning products on DPO's that include reports on Macroeconomic Framework of DPOs and the links of DPOs to Bank's Public Expenditure Reviews.

² OPCS Good Practice Note

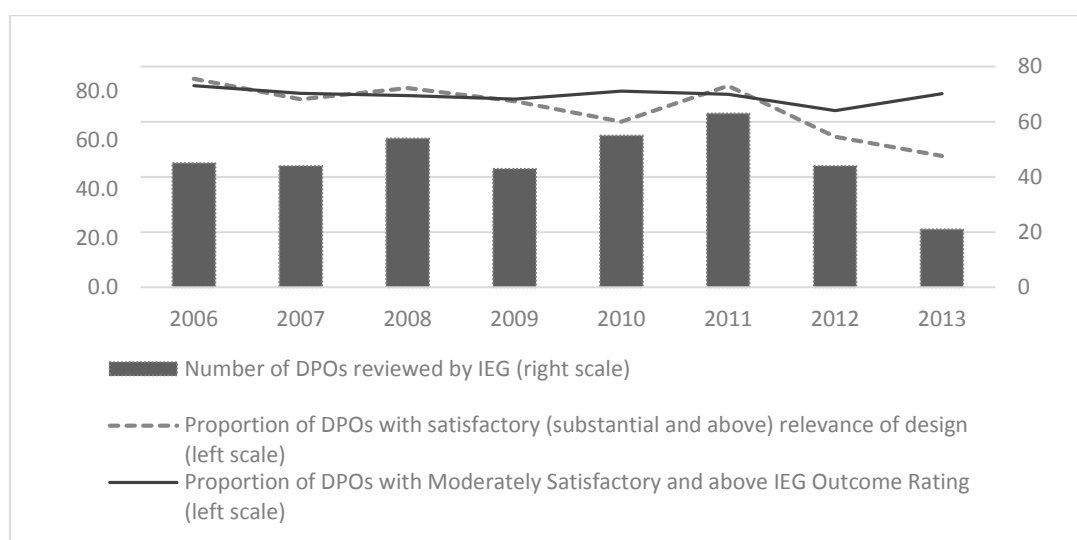
³ 2012 DPO Retrospective

1.INTRODUCTION

their contribution to the results chain for DPOs closed in 2009–2011⁴. According to that review, 80 percent of DPOs had satisfactory RFs, which was an improvement over the previous Retrospective periods⁵.

1.5 IEG’s validation and evaluation largely draws on the quality of the RF for its assessment of relevance of design of DPOs. Other aspects of the RF are captured through assessment of the quality of the M&E framework in IEG’s Implementation Completion and Results Report Reviews (ICRRs). Figure 1.1 shows the ratio of DPOs with satisfactory relevance of design ratings (substantial and above), which is one input in defining the overall IEG outcome rating. In addition to assessing the RFs for ICRRs, IEG has looked at various aspects of RFs for DPOs as part of the 2010 evaluation of Poverty Reduction Support Credits. This review aims to contribute to the Bank’s knowledge of the RFs for DPOs.

Figure 1-1. Outcome Rating and Relevance of Design Ratings for DPOs⁶



1.6 The structure of the review is: chapter 1 presents the methodology; chapter 2 has the findings with respect to the presentation and clarity of RFs; chapter 3 discusses evidence of the quality of PAs and their contribution to the robustness of RFs; chapter 4 discusses findings with respect to the use of programmatic approach in DPOs relevant for RFs; chapter 5 discusses strengths and weaknesses in the M&E framework of DPOs; chapter 6 discusses reporting of DPO results at completion and implications for IEG validation and evaluation of DPOs.

⁴ 2012 DPO Retrospective

⁵ This ratio declines if the threshold of prior actions is changed.

⁶ Some Implementation Completion and Results Reports of DPOs exited in FY12–13 were not reviewed by the Independent Evaluation Group at the time of preparation of this report.

2. Methodology

2.1 This assessment builds on evaluative findings from previous work by the Independent Evaluation Group (IEG), including Implementation Completion and Results Report Reviews, Poverty Reduction Support Credit (PRSC) and crisis response thematic evaluations, and Project Performance Assessment Reports of long-term programmatic series (Tanzania and Vietnam PRSCs). The review also draws on 14 desk reviews of development policy operations (DPOs) conducted specifically for the series. Additionally, the assessment draws on a review of randomly selected program documents and Implementation Completion and Results Reports (ICRs) for more information on various aspects of Results Frameworks (RFs).

2.2 Table 2.1 presents the coverage and methodology used in addressing these evaluative questions:

- Do RFs present a clear statement of program objectives and outcomes? Do the prior actions have sufficient institutional depth and criticality to achieve the targeted results?
- What are the implications of the programmatic approach for RFs?
- Does the monitoring and evaluation framework capture the impact of the DPO on country outcomes?
- Do ICRs provide a comprehensive account of attribution of the Bank's inputs to programs results?

2.3 This review also documents recent changes in the Bank's approach to RFs of DPOs, examines their impact on the quality of the instrument, and aims to derive lessons for IEG in improving validation and evaluation of DPOs.

2.4 The framework for assessing prior actions (PAs) was based on: (i) the quality criteria for policy actions identified in the guidance note from Operations Policy and Country Services on results in DPOs⁷; and (ii) strengths and weaknesses observed in various IEG evaluative exercises. The review evaluates the robustness of RFs based on how each PA meets five criteria: (i) relevance, (ii) criticality, (iii) measurability, (iv) additionality, and (v) efficacy⁸ (see chapter 4 for more information).

⁷ Good Practice Note for Development Policy Lending: Results in Development Policy Lending OPCS 2011

⁸ These criteria are further elaborated in section 4 on Prior Actions.

2. METHODOLOGY

Table 2-1. Coverage and Methodology Used in Evaluative Questions

Section	Methodology
Presentation and clarity of RF	This section draws on evaluative evidence from DPO PPARs and ICRRs prepared in 2012–2014; 14 case studies; and review of 60 randomly selected program documents of DPOs approved in FY12–14 (20 operations per each FY, representing 55, 70, and 65 percent of all new DPOs in FY12, FY13, and FY14, respectively).
Quality of prior actions	This section draws on evaluative evidence from DPO PPARs and ICRRs from 2012–2014 and 14 case studies. Systematic assessment of the quality of PAs based on five characteristics (i.e., relevance, criticality, additionality, measurability, and efficacy) is derived from two PPARs and 14 case studies. This section also draws on a review of program documents (42) and ICRs (26, wherever available) of 42 programs that included specific type of policy actions of interest for this report.
Quality of RF in programmatic approach	This section draws on evaluative evidence from DPO PPARs and ICRRs from 2012–2014 and 14 case studies
Monitoring and Evaluation Framework	This section draws on evaluative evidence from DPO PPARs and ICRRs from 2012–2014; 14 case studies; review of all IEG ICRRs on DPOs prepared in 2012–2014; and review of all ICRs closed in FY12, FY13, and FY14.
Reporting of results in ICRs	This section draws on evaluative evidence from DPO PPARs and ICRRs from 2012–2014; 14 case studies; and review of 26 ICRs selected for analysis for the section on PAs.
IEG validation and project level evaluation	This section draws on a review of all IEG ICRRs on DPOs prepared in 2012–2014.

Note: DPO = development policy operation; FY = fiscal year; ICR = Implementation Completion and Results Report; ICRR = ICR Review; IEG = Independent Evaluation Group; PA = prior action; PPAR = Project Performance Assessment Report; RF = Results Framework.

3. Presentation of Results Frameworks

3.1 The Results Frameworks (RFs) for a development policy operation (DPO) is presented in its policy matrix. The matrix should include the objectives, critical actions supported by the Bank, and outcomes and indicators (with baseline and target values) to measure program objectives. The 2011 guidance note from Operations Policy and Country Services (OPCS) presents two models for policy matrices—for stand-alone operations and for a programmatic series (World Bank 2011). Since FY14, OPCS has promoted a new model of policy matrix that has increasingly applied to newly adopted DPOs. This chapter is based on analysis of recent changes in OPCS guidelines, a review of program documents of 60 relatively new DPOs approved in FY12–14 (20 DPOs randomly selected for each year, covering more than 50 percent of all operations), and evidence from previous IEG evaluations.

Overview of Recent Changes in OPCS Guidelines

3.2 In 2009 and 2011, OPCS’s good practice guidance notes introduced some changes in the presentation of DPO RFs. DPOs approved since then have over time followed these guidelines. Since FY14, operations are increasingly reflecting the new, more simplified OPCS template for policy matrices.

3.3 Elimination of nonprior action policy actions (“benchmarks”) from the policy matrix is the main change. The use of benchmarks dates back to adjustment lending operations when the key distinction between benchmarks and prior actions (PAs) was their timing—PAs were implemented before programs went to the Board while benchmarks were expected to be implemented before completion. Benchmarks lost their importance as a performance monitoring tool in the mid-2000s with the introduction of explicit programmatic series and indicative triggers and more systematic use of quantitative monitoring and evaluation (M&E) indicators that transformed the benchmarks into nonbinding policy actions. The use of the term benchmark to describe these nonbinding actions became controversial. This review examined 60 DPOs approved in FY12–14 (more than 50 percent of all new DPOs approved). Only five included benchmarks, and DPOs approved in FY14 did not include any.

3.4 According to OPCS, the rationale for not including benchmarks in the policy matrix is to avoid the impression of disguised conditionality⁹ (World Bank 2011, 1). IEG’s review of evidence indicates that the use of benchmarks has been inconsistent in recent years. While their inclusion often augmented the policy content of DPOs at least in appearance, most of the program documents and the Implementation Completion

⁹ OPCS 2011 Good Practice guidance note, page1

3. PRESENTATION OF RESULTS FRAMEWORK

and Results Reports (ICRs) provided only selective information by reporting good cases of implementation and neglecting cases of poor performance. To enhance accountability, it was appropriate to drop actions from the policy matrix that were not monitored by the Bank and had uncertain attribution to Bank's interventions.

3.5 Presentation of indicative triggers in the policy matrix also has changed. Prior to 2011, a typical program document of the first operation in a series of three or more DPOs presented indicative targets only for the next operation. The reform program for outer operations was presented vaguely, usually without a distinction between benchmarks and triggers. Now the policy matrix of the first operation is required to present the reform program covering all proposed operations in the series. All reviewed programmatic series approved in FY12–14 followed this approach. The approach of presenting a complete program at the onset, with indicative triggers covering *all* proposed operations (that can still be flexible) is associated with stronger RFs. There is evidence that the series with vague presentation of triggers for outer operations were compromised by excessive flexibility (see chapter 5).

3.6 Presentation of RFs in the policy matrix was simplified. The new OPCS policy matrix template introduced in FY14 (but not incorporated in the good practice guidance note as of preparation of this report) only contains PAs and outcome indicators. (The strengths and weaknesses of this approach are discussed in following sections.)

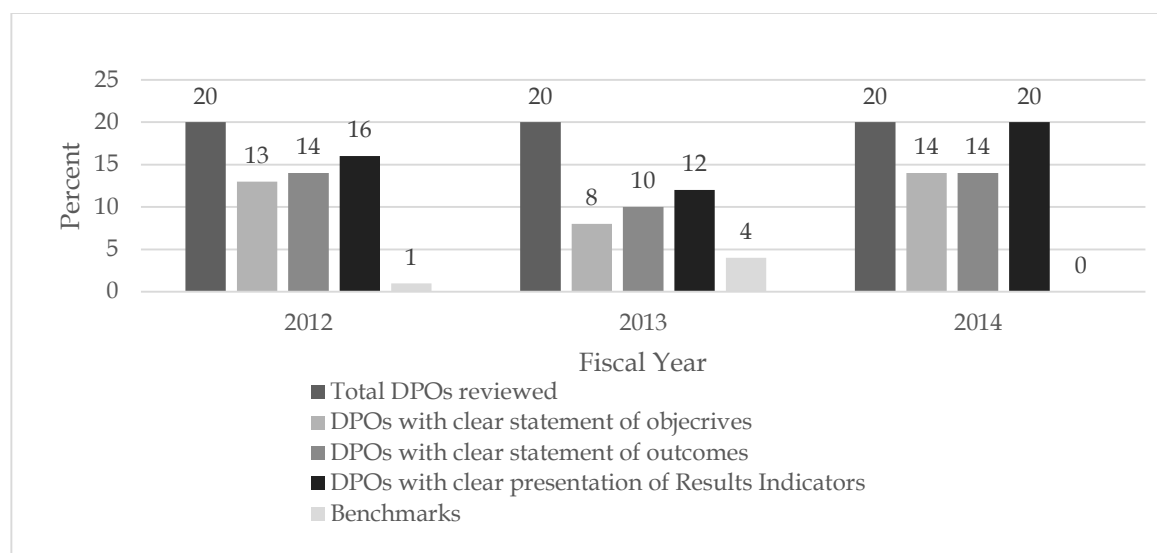
3.7 Other changes in DPOs include: (i) most of the DPOs no longer use intermediate outcome indicators; (ii) there is no longer a distinction between results indicators and Program Development Objective indicators; (iii) the M&E framework is no longer presented separately from the policy matrix; (iv) the number of DPOs in one series is now capped at three in most cases (series of four to five operations were not uncommon in the mid- to late 2000s); and (v) the time allocated to ICR preparation was changed from six months to 12 months in FY15. The report discusses most of these changes in light of evidence in respective chapters.

Issues with the Results Framework Presentation

3.8 Recent changes have helped to enhance the clarity of RFs presentation, but there are shortcomings. The statement of objectives remains a challenge in DPOs. In many reviewed, the objective was stated differently in the program summaries and policy matrices. Financing agreements of DPOs, which do not include a statement of objectives, may also contribute to the problem. Many programs approved before FY14 had different statements of objectives in the program summary and policy matrix as the 2011 OPCS template allowed multiple objectives per policy area. The new template in use since FY14 addresses this issue to some extent. However, several recently approved

DPOs indicate confusion between objectives, pillars, and policy areas¹⁰. As shown in figure 3.1, the statement of objectives continues to remain a problem.

Figure 3-1. Presentation of Results Framework¹¹ in DPOs Policy Documents



3.9 Clarity of statements of objectives is very important for IEG evaluation. ICRs and ICR Reviews (ICRRs) tend to stick to broader definitions of objectives usually from the “program summary” section of program documents, rendering the statement of subobjectives in a policy matrix useless in most cases. This would be less of a problem for most DPOs approved in FY14 and after, as the new template addresses this issue. However, this review finds that identifying objectives in some programs based on the new template is difficult as some of them no longer provide statements of objectives in the policy matrix, and in many cases concepts such as “objectives,” “pillars” and “policy areas” are used interchangeably.

3.10 Many DPOs do not contain a statement of outcomes. According to the OPCS guidelines, DPO policy matrices should include outcomes. Outcomes are an important part of the results chain in DPOs linking the PAs supported by the Bank to DPOs’ objectives. In many cases, however, instead of an explicit statement of outcome, policy matrix contains only result indicators. RFs suffer from this simplification in cases when

¹⁰ See, for example, Kiribati 2014, Nigeria 2014

¹¹ A clear statement of *objectives* means that the objectives are stated the same in different parts of the report, and the policy matrix includes a clear statement of objectives. The statement of *outcomes* would be unclear if for more than two prior actions there is no explicit statement of outcome while results indicator is output oriented. Clear presentation of *results indicators* means that the results indicators presented in the program summary and policy matrix are the same. Randomly selected DPO’s includes programs that were part of the programmatic series only if they were the first in the series.

3. PRESENTATION OF RESULTS FRAMEWORK

these indicators are output rather than outcome oriented. This effectively leads to a disappearance of the outcome from the results framework thus affecting the robustness of RF and its evaluability (see box 3.1 for examples). In 60 new DPOs approved in FY12–14 this review found problems with statement of outcomes in 22 of them¹². However there were also good examples of clear distinction between outcomes and result indicators that contributes to the quality of RF¹³ (see appendix B).

Box 3.1. Examples of Lack of Explicit Statements of Outcomes that Undermine the Clarity of the Results Framework

Examples when outcome indicators are inadequate substitutes of missing	Observations
Revenue of Naftogaz from sales of gas to residential consumers: Baseline (2013): 6.9 billion Target (2014): 10.4 billion	Based on the program document, the implicit outcome is to eliminate tariff setting below cost-recovery levels, which will contribute to the program's objective of improved public sector governance. The outcome indicator (revenues of utility company to increase from 6.9 billion units to 10.4 billion units of national currency in one year) potentially can measure the extent of achievement of expected outcome. In that respect, the choice of outcome indicator may be appropriate. But this indicator cannot fill the gap of a missing explicit statement of outcome (improved tariff setting). The reason is, at completion and reporting, this indicator may not be considered still relevant, and more evidence may need to be found. This may happen if, for example, there is a significant price change from an exogenous shock (e.g. exchange rate depreciation). This may lead to an increase in the utility's revenues that is neutral for its financial health, because of lack of progress with tariff setting mechanisms. Stating a clear strategic outcome, in this case "eliminating tariff setting below cost-recovery levels," would allow for comprehensive assessment of achievement of outcomes, including using the proposed outcome indicator and other evidence, if needed.
Outcome Indicator: Total budget assigned to finance provision of specific measures to protect women who are victims of gender violence, as percent of total budget expenditure: Baseline (2012) = 0 Target (2014) = 0.01	This outcome indicator is a good choice for monitoring and evaluation of an implicit program outcome (under the objective of gender equity), which is the provision of publicly funded services to victims of gender violence (based on the program document). This outcome indicator cannot fully fill the gap of a missing explicit statement of outcome as it may not be still relevant at completion. This could happen if the funding is made available (target met), but delivery was hampered by inefficiencies, and no service is actually delivered to the victims. By establishing a separate explicit outcome in the policy matrix, it would become possible to evaluate achievement of outcomes using alternative evidence, if needed

3.11 Lack of an explicit statement of outcomes creates several issues. First, the lack of an explicit statement of outcome creates a situation in which program outcomes have to be derived from results indicators instead of the other way around. Second, in a programmatic series, results indicators may change, be dropped, or added, which

¹² Defined as two or more outcomes not stated explicitly and clearly in the RF (mostly cases when no distinction is made between outcome indicators and outcomes, while the indicators are output oriented and cannot substitute outcomes).

¹³ Paraguay 2012, Colombia 2012, Myanmar 2013, and Grenada 2014 are among "good practice" examples.

would technically change the program's expected outcome – a situation to avoid. In cases reviewed by IEG, one-third of results indicators were dropped at some point, effectively leading to the disappearance of many expected outcomes. Third, program outcomes are part of the results chain, and their adequacy is evaluated as part of relevance of design. Choice of the outcome indicators and their target values are issues belonging to the relevance of M&E. Not stating the implicit outcome in the policy matrix of DPOs in cases when it is necessary affects IEG's evaluation of those DPOs. Box 3.2 presents a model of policy matrix based on good practice DPOs identified by this review.

Box 3.2. A Model of Best Practice Results Framework Presentation in the Policy Matrix of Development Policy Operations

Many DPOs reviewed provide good examples of Results Framework presentation. They share similar characteristics such as clear statements of objectives (in both the program summary and policy matrix) and outcomes. For illustration purposes, the figure below shows a summary model of a policy matrix that best describes the result chains underpinning DPOs based on the Bank's approach of streamlining and simplicity.

Objective 1			
Policy Area 1 (if needed)			
Prior Action 1	Outcome 1	Indicator 1	
		Baseline	Target
Prior Action 2	Outcome 2	Indicator 2	
		Baseline	Target
Policy Area 2			
Prior Action 3	Outcome 3	Indicator 3	
		Baseline	Target
Objective 2			

3.12 Presentation of medium term policies and expected outcomes of the government program in policy matrices of stand-alone operations might be ineffective. To address the problem of short time horizons of stand-alone operations, the 2011 OPCS guidance note recommended incorporating information on a government's medium term program, noting however that this is not part of the DPO supported program (World Bank 2011). The latest template of the policy matrix promoted by OPCS no longer uses this approach for stand-alone operations. This review finds that this approach did not add to the robustness of RFs, especially in light of recent changes in the timing of the ICRs of DPOs that was extended to 12 months. This option can be inappropriately used to mask output rather than outcome orientation in some DPOs. As a general note, this review finds that, in many cases, citing relatively short time horizon of stand-alone DPOs is an inadequate explanation

3. PRESENTATION OF RESULTS FRAMEWORK

for not pursuing tangible policy change. Even stand-alone DPOs can span up to two years (from the Project Information Document to the ICR) allowing for the incorporation of meaningful medium term policy targets as part of DPOs.

Presentation in Legal Documents Underpinning DPOs

3.13 Legal underpinning of DPOs and RFs is different from investment lending. Legal agreements of Bank DPOs do not include a statement of objectives, which differentiates it from investment lending. PAs are the only element of RFs included in the legal agreement, which is also different from the investment lending approach in which forward looking elements of policy components are included in the legal agreement. This study could not find evaluative evidence on the implications of this approach for incentives and results orientation in DPOs since there was no previous study covering this aspect. Because some stakeholders felt strongly that this feature of DPOs may weaken programs, this may merit further scrutiny.

4. Quality of Prior Actions

4.1 Prior actions (PAs) are the key elements of the Results Frameworks (RFs) for development policy operations (DPOs). They are the only elements of RF reflected in the legal agreement between the World Bank and borrowers. They are expected to trigger policy and institutional change to an extent that contributes to the country's potential for equitable growth. PAs are expected to draw from the government program, but also reflect the Bank's contribution to the reform agenda through policy dialogue and technical assistance to ensure added value. All components of results chains are strongly linked to the PAs, allowing them to serve as a stepping stone for assessing the overall RF.

4.2 Assessment of prior actions is challenging and context specific. The Independent Evaluation Group (IEG) reflects on the quality of prior action in its evaluation and validation work to various extents. As part of this review, evidence on the quality of PAs from Implementation Completion and Results Report Reviews (ICRRs), Project Performance Assessment Reports (PPARs), and case studies was systematically identified, synthesized, and categorized based on these characteristics: (i) relevance, (ii) criticality, (iii) additionality or value added, (iv) measurability or monitorability, and (v) efficacy (table 4.1).

Table 4-1. IEG Criteria for Assessing the Prior Actions of Development Policy Operations

Criteria	What Does IEG Look for?
Relevance	The extent to which PAs are relevant to objectives/outcomes, and whether there are links with actions from both previous and subsequent operations in case of programmatic series.
Criticality	The extent to which PAs have sufficient institutional depth to trigger policy/institutional change. Policy actions that are excessively process-oriented, easily reversible or only indicate intentions should be avoided.
Additionality	The extent to which PAs reflect the World Bank's value added with respect to the borrower's reform agenda. This is assessed based on factors such as the timing of the action, cross-conditionality with IMF, and Bank engagement in reform areas.
Measurability	The extent to which the expected impact of PAs is measurable. This largely depends on the quality of the M&E framework and the links between results indicators and PAs.
Efficacy (completeness)	The extent to which PAs become completed as an output for the results chain (different from "efficacy" used in ICRR which refers to achievement of outcomes). This refers mostly to prior actions such as draft laws, decrees, and strategies that require follow-up steps to be completed.

Note: DPO = development policy operation; ICRR = Implementation Completion and Results Report Review; IEG = Independent Evaluation Group; IMF = International Monetary Fund; M&E = monitoring and evaluation; PAs = prior actions.

4. QUALITY OF PRIOR ACTIONS

4.3 The Operations Policy and Country Services (OPCS) Good Practice Note provides guidance on prior actions (World Bank 2011). Relevance, criticality, and monitorability are the key characteristics explicitly discussed in the note, and they are used for assessment of PAs for the retrospective by OPCS. IEG's approach extends this method further based on availability of project-based evidence as part of its evaluations. However, unlike OPCS's assessment for its retrospective, this review does not provide assessment of all PAs. It only reflects on the evaluative evidence of the quality of PAs from various IEG products.

Relevance

4.4 Prior actions should necessarily be relevant with respect to both program level objectives and to specific results and outcomes defined in the policy matrix. In cases of a programmatic series, the relevance of PAs should also be viewed in light of the links with actions from previous and subsequent operations.

4.5 Very few programs had prior actions that were not relevant to objectives; however, the relevance with respect to the results appears to be more challenging. For example, a prior action on adopting the law on HIV/ AIDS under the objective of improving health care may be quite relevant, while the links to the identified results (i.e., child malnutrition, poor people with free health insurance, and use of modern health care facilities by the poor) is less obvious in a country with a very low prevalence of HIV/ AIDS. This inconsistency was present especially in some Poverty Reduction Support Credit (PRSC) type, multi-sectoral operations.

4.6 Shortcomings in presentation of RFs may affect relevance. The confusion of outcomes with results indicators (discussed in section 3) – dropping a result indicator or replacing it by another – may change the statements of results affecting the relevance of PAs with respect to results or outcomes. This can be fixed by improving the presentation of RFs to clearly distinguish outcomes from results indicators.

4.7 In some multi-sectoral operations, the links between prior actions in the same policy areas of individual DPOs in the series may be weak. This occurred mostly in cases where the reform agenda was not fully defined at the onset and presented in the program document of the first operation¹⁴.

Criticality

4.8 Criticality of prior actions is important for triggering lasting policy and institutional change. Criticality is well defined in the OPCS guidance note: policy

¹⁴ Vietnam and Benin PRSCs

actions that are excessively process-oriented, easily reversible, or only indicate intentions of the government as opposed to actual substantive changes should be avoided. OPCS uses an implicitly normative classification of PAs for its retrospective to capture the extent of criticality (see box 4.1).

Box 4.1. Operations Policy and Country Services' Classification of Prior Actions

Operations Policy and Country Services (OPCS) distinguishes between three types of prior actions (PAs): (i) policy and institutional reform focused actions; (ii) process-focused actions; and (iii) implementation-focused actions (World Bank 2013a). Recognizing that different context requires different actions, OPCS's approach implicitly suggests that policy and institutional actions are more critical than process and implementation actions. According to this classification, the ratio of policy and institutional actions covered by the 2012 retrospective varied from 40 percent to 60 percent of all actions across Regions. According to the approach used in this review, some policy and institutional actions may be noncritical if they do not represent sufficient institutional depth (for example, draft policy documents not vetted and not approved by the government).

Source: World Bank (2013a).

4.9 IEG's experience with DPOs indicates that, for prior actions to be critical, there should be sufficient institutional depth and a certain degree of consensus among various stakeholders to avoid policy reversals. However, criticality can mean different things in different contexts. In some cases, a law may be necessary for an institutional change while, in other cases, government decrees and even lower level actions that complete a reform process can be critical for a program. IEG's review of criticality of PAs has some advantages because its project level evaluations allow judgments based on specific context.

4.10 Existing evaluative evidence suggest that many DPOs include policy actions that are not critical in the given context. In 14 desk reviews and IEG PPARs, the number of noncritical actions and their share in total PAs varied substantially – from 3 out of 31 PAs (Peru) to 8 out of 51 (Vietnam PRSC) for a programmatic series, and from 0 out of 7 (Albania) to 3 out of 8 (Côte d'Ivoire) in stand-alone operations. Interpretation of these numbers, however, is not straightforward. In some cases, DPO policy matrices may include a number of transformative actions (that can ensure a satisfactory outcome) along with a few actions, which lack criticality. This is observed in cases of sector-specific operations (e.g. Latvia). Based on this, one may suggest that the quality of policy matrices should be defined by the presence of necessary and critical actions rather than by the fact that policy matrices include some noncritical actions. However, the use of less critical PAs in otherwise effective interventions may create a precedent for PAs of lower quality.

4. QUALITY OF PRIOR ACTIONS

4.11 The key challenge to criticality is the level of institutional depth. In many cases, the PAs are for actions undertaken at the agency level or below and lack institutional depth. Examples of such PAs are assessments, evaluations, and reports prepared by various agencies and shared with the World Bank that do not have important regulatory or operational implications. Although such actions may address important issues, they don't necessarily reflect actual commitment or intention at the sufficiently high level necessary for implementing reforms. Some Implementation Completion and Results Reports (ICRs) do not provide full account of such actions. In most cases, it remains unclear what exactly was accomplished through these actions.

4.12 Incorporation of draft laws as PAs before government approval is controversial. This review found that, in 2009–2013, 15 operations dealt with submitting a draft law by government agencies (ministries or below) for review by the cabinet. Box 4.2 provides more information on PAs on laws. The rationale of this approach is questionable as governments should be able to adopt draft laws and submit them to the parliament if there is strong reform momentum. The review additionally identified 17 programs that dealt with approving draft laws by the government without submitting to parliament (see more information in the Efficacy section of this chapter). In some cases the ICRs or subsequent program documents did not provide a status report on those actions.

Box 4.2. Use of Prior Actions on Laws

Cases of draft laws used as prior actions (PAs) may be of special evaluative importance. First, laws are considered critical policy actions that can trigger policy and institutional change. Second, most Implementation Completion and Results Reports (ICRs) report on the status of actions on law, while ICR follow-up on other actions is less systematic.

According to World Bank guidelines, submission of draft laws to nations' parliaments is an adequate PA. Adoption of laws as a PA is not encouraged as it is considered beyond the government's power. The table below shows that in fact the Bank uses various types of actions on laws as PAs. In 11 programs, the Bank used adoption of laws by parliaments as PAs, while in 32 cases the Bank used draft laws before submission to the parliament as PAs. The highest number of PAs is on submission of the draft laws to parliament. There is evidence that the use of draft laws at an early stage is controversial (see section in this chapter on Efficacy).

Number of Development Policy Operations Using Laws as Prior Actions

FY	2010	2011	2012	2013	2014	Total
Adopted laws	5	2	2	1	1	11
Draft laws submitted to parliament	15	8	20	19	2	64
Draft laws approved by government but not submitted to parliaments	7	6	0	2		
Laws drafted but not vetted by the government	1	5	3	1	5	15
Total prior actions on laws	28	21	25	23	10	107

Source: OPCS Database on Prior Actions

4.13 Adopting various draft documents that are under government authority is also controversial. If reform momentum is present, governments should be able to adopt required regulations that are under their authority, such as decrees, decisions, master plans, strategies, and policy frameworks. This review identified 26 programs for the period 2009–2013 that included PAs on preparing such draft documents at the level of ministries and below (e.g., draft action plans, draft decrees, draft policy papers, trials, or below levels). This suggests that the consensus within respective governments might have been lacking because these documents could have been approved to enhance the institutional content of PAs supported by DPOs. In most cases reviewed, no strong evidence was found in program documents or ICRs on follow-up on these actions, including their final adoption.

4.14 Despite OPCS’s explicit recommendation in the practice note, statements of intentions are still used as PAs. In three out of 14 case studies there were statements about reform intentions made mostly at the ministerial level used as PAs. Those include letters issued specifically to the Bank or other international organizations stating government commitment to reforms. Another set of noncritical PAs refer to actions on the publication of certain documents and strategies in an environment where transparency does not seem to be an issue. Many process oriented PAs, such as implementation of an ongoing strategy or annual budget, were, in most of the cases, found lacking criticality unless they added certain important dimensions to implementation.

Additionality

4.15 The concept of additionality, or the value added of PAs, covers several aspects. First, it assesses the extent to which actions added value as the result of the Bank’s engagement in a given reform area. The OPCS guidance note on designing DPOs suggests “what would be discouraged is a policy matrix which simply documents government reforms¹⁵ without any inputs from Bank staff, for instance in order to give credit to the government for carrying out certain reforms” (World Bank 2011, 4). When PAs were implemented before the Bank’s engagement through DPOs, and there is no evidence of a Bank-client cooperation in respective specific reform area, it is clear the action would have taken place without the Bank’s support. In this case DPOs lack value added in terms of their contribution to policy and institutional change. Policy actions implemented before the Bank’s decision to do a DPO and without the Bank’s

¹⁵ Page 4

4. QUALITY OF PRIOR ACTIONS

contribution are seen mostly in crisis or emergency support and in stand-alone operations. The Bank can use other simplified options of budget support for such cases.

4.16 Other situations of additionality clearly emerging as an issue includes cases of cross-conditionality with the International Monetary Fund, and cases when the same prior action is used a number of times for recurrent events. IEG has identified cases when the PAs of DPOs are the same as the IMF's structural benchmarks (e.g., Honduras, Latvia). PAs of various budgetary functions of a recurrent character, such as submission of a draft budget law to the parliament every year, approval of annual budget execution reports can be used a number of times in subsequent operations (Côte d'Ivoire, Indonesia). DPOs should in fact aim for institutionalization of government functions, and incorporation of those actions in DPOs is more controversial in countries with relatively high policy and implementation capacity. If actions of this type are critical for continuing engagement through DPOs in particular cases, the Bank can monitor these developments as part of ongoing policy dialogue and monitoring of risks.

Measurability

4.17 Measurability or monitorability is an important feature in a results-focused framework. This is highlighted as one of the key concepts in OPCS's good practice note on DPO design. A good framework should be able to assess the scope of change as a result of Bank supported policy actions. Key challenges of measurability of PAs include lack of results indicators in M&E frameworks that are linked to the prior action, failure to report data on result indicators, and the dropping of result indicators altogether. In some cases, the expected results go beyond the timeline of the DPO and, therefore, are not reported in the ICR. For example, an adopted law may be designed to actually enter into force in a number of years after adoption; thus, despite the importance of such an action, the M&E framework would not be able to capture the results (e.g., Latvia). In some situations the result indicator and the PAs could be the same, which is strongly discouraged by OPCS (e.g., Côte d'Ivoire). Most issues related to measurability of PAs are discussed in Section 6 on M&E.

Efficacy

4.18 IEG has introduced the concept of efficacy of prior action for this study. It is different from IEG's assessment of a project's efficacy that is used in project validation and evaluation, which refers to the final outcome of intervention. In this context, efficacy refers to the degree to which PAs become completed as an output for the results chain. For example, a law that is to be passed by parliament as a prior action needs to be enacted, and, if it does not happen, the prior action does not constitute a completed

output. Another example would be a draft national strategy as a prior action that is expected to become operational and lead to desired policy changes. Underperformance of a results indicator would not necessarily indicate lack of efficacy of respective prior action since other factors may be involved. However, large deviation from targets or a reversal in direction of expected changes might be a signal of possible problems in the quality of PAs.

4.19 IEG’s review of evaluative evidence indicates that most PAs that lack criticality did not lead to the change necessary for achieving outcomes. For example, the Mid-Term Expenditure Framework developed as a ministry-level exercise and not embedded in formal budgetary processes did not help with the introduction of medium term budgeting in Côte d’Ivoire. At the same time, some critical actions may lack efficacy because the environment may change, and assumptions may turn out to be wrong. For example, adoption of a law on the re-introduction of the second pillar of a pension system was, by all accounts, a critical action requiring national consensus building; however, this lacked efficacy because its enacting was delayed after the closing of the DPO as a result of unexpectedly poor fiscal conditions¹⁶.

4.20 Many PAs dealing with the submission of draft legislation by governments to parliaments lack efficacy. Delays in adoption by the parliament are widespread. An example is the Law on Competition in Ghana that was submitted to the parliament in 2009 as a prior action but was approved in February 2015. The use of such PAs may be especially problematic in countries where the transaction and political costs of submitting draft legislation to the legislature is relatively low. This allows authorities to submit nonviable drafts as an implementation of PAs under DPOs.

4.21 The efficacy of PAs on draft documents prepared at the government level and not submitted to parliament is largely lacking. Table 4.1 provides information on PAs for draft laws before submission to the parliaments (approved by the government but not submitted, or a draft prepared by a ministry or lower level agency and submitted to the government). Out of around 25 programs with such PAs, evidence was available through ICRs for nine of them, of which seven reported no progress at completion.

4.22 PAs that cover a relatively small subset of agencies where a change should take place often lack efficacy. This is especially the case of PAs on pilot initiatives. For example, PAs such as the introduction of a new public financial management model in three out of 61 provinces, in some cases would lack efficacy if there is no clearly defined scaling-up strategy and commitment.

¹⁶ Latvia Social Reform Program

4. QUALITY OF PRIOR ACTIONS

Table 4-2. Status of Draft Laws Not Submitted to the Parliament used as PAs

Progress	No Progress
Guinea Bissau: economic governance	Iraq: fiscal sustainability
Uruguay: programmatic reform implementation	Jordan: recovery under global uncertainty
	FYR Macedonia: development policy loan
	Morocco: sustainable access to finance
	Nepal: Financial Sector Stability Credit
	Tanzania: Poverty Reduction Support Credits
	Togo: economic recovery and governance

Source: ICRs of Respective DPOs.

Other Considerations in Quality of Prior Actions

4.23 This review has also looked at the issues of missed actions, which is different from the question of the scope of the program. In this particular context it refers to actions in areas that are explicitly or implicitly presented in the program document as critical for achieving the program's objectives but not incorporated in the policy matrix. An example is the Nigeria Financial Sector DPO where the focus of PAs is on banking sector liquidity but actions on solvency (a major challenge identified in the program document) were missing. The program document implicitly suggests that progress in this area would be impossible without addressing structural issues in the banking sector raising concerns of the RF and program efficiency. Another example is the Thailand Public Sector Development Loan, which focuses on public financial management and service delivery but fails to address public investment management, an area identified in the program document as one of most challenging areas in need of reform in the public sector.

4.24 The quality of PAs in multi-sectoral operations is very critical for a program's results orientation. Both multi-sectoral and sector specific DPOs can only have a certain number of PAs (about 10). In some multi-sectoral, programmatic series, not all policy areas get to have a prior action in each operation. This puts a high premium on the quality of PAs of multi-sectoral programs. In a typical multi-sectoral program with three to four mostly unrelated policy areas, the PAs should be carefully selected and promoted by the Bank to ensure that they truly contribute to the desired institutional change. Interestingly, this review does not find that PAs in multi-sectoral operations are more critical than those of sector programs if viewed in isolation. In fact, PAs in sector specific operations benefit from better alignment with the results and better complementarity with other PAs. This issue relates somewhat to the issue of depth versus breadth of DPOs. IEG's evaluative evidence indicates that programs with emphasis on breadth achieve major results only with the support of transformative

policy actions (e.g., Vietnam's first series of PRSCs implemented in 2001–2006 versus the second series implemented in 2007–2012).

Summary on Quality of Prior Actions

Key points on the quality of PAs are the following:

- PAs are drawn from government programs and, in that respect, they reflect the depth and ambitions of the borrowers' reform agendas. To improve the results of the reforms through DPOs, the Bank should raise the bar on the quality of policy actions that DPOs support.
- There is evidence that the Bank's approach to PAs has been improving. Although this review does not provide an assessment of the quality of PAs across time, it notes that, in a number of aspects, the frequency of noted problems appears to be less with more recent operations, especially in the area of monitorability of PAs.
- Further increasing corporate standards on the quality of PAs, possibly through more explicit guidance to discourage DPO teams from using PAs of certain types that lack functionality, will empower Bank teams to identify and negotiate more critical policy actions with the borrowers.

5. Results Frameworks in Programmatic DPOs

5.1 The programmatic approach to development policy lending is the Bank's preferred option when possible¹⁷. Programmatic development policy operations (DPOs) are typically implemented in countries with well-defined, medium term reform programs, including many Poverty Reduction Support Credit (PRSC) countries that have sufficient institutional capacity to develop and commit to a longer term reform program. Out of about 300 DPOs that closed in FY09–13, about half belonged to 65 different programmatic series (2.5 operations per series on average).

5.2 The programmatic approach allows the Bank to extend its interventions up to five or six years. It also allows for sequencing of the reform program and greater flexibility through the use of the triggers. As the triggers are only indicative, this approach is different from the multi-tranche approach used by the Bank where tranches were established at the preparation stage and were rigid. This review by the Independent Evaluation Group (IEG) draws on project-level evaluations and special desk reviews of several programmatic series (Benin, Indonesia, Latvia, Peru, Tanzania, and Vietnam) as well as evaluative evidence available from IEG validation of Implementation Completion and Results Reports (ICRs).

5.3 The Bank's approach to the presentation of the Results Frameworks (RFs) in a programmatic series has changed. There is more clarity with respect to the indicative triggers (discussed in chapter 2). All series approved in FY12 or beyond have triggers presented for all operations at the outset of the series, whereas in older programs the triggers were fully specified in the program documents only for the next operation in the series. IEG's evaluative evidence suggests that this approach improves the robustness of the results chain while also allowing the flexibility necessary for program implementation. The current model of presentation of the RFs allows for the incorporation of medium term results indicators if necessary.

5.4 Programmatic series allow flexibility in a number of ways. Flexibility normally means adjusting the ambition, directions, and emphasis of reforms throughout the progression of the series and allowing incorporation of reforms that might not have been envisioned at the preparation stage. In some rare cases, the RFs may be modified to accommodate an additional operation that was not planned at the preparation stage (Indonesia). The RFs absorb the changes by modifying the triggers that become PAs, dropping the triggers for subsequent operations (they don't become PAs), adopting

¹⁷ According to the OPCS guidance note, "Under normal circumstances, a standalone single tranche operation would not be considered the most appropriate structure given that most policy and institutional reforms are medium- to long-term processes of multiple sequential actions" (World Bank 2011, 12).

partially met triggers as PAs, and introducing new PAs outside the pool of the triggers. The programmatic approach also allows revisions in the results and results indicators based on program performance, including changes in target values and dropping or adding results indicators.

5.5 This study collected evidence on the use of the programmatic approach and its implications for robustness of the RFs of DPOs. The programmatic approach may be used for raising the ambitions of the reforms as series progresses. Among the series that were reviewed for this study, only in one case did the actual reform program go somewhat beyond the initial plan in terms of the program's depth by adopting more ambitious reforms (i.e., Peru). In most cases, the programmatic approach helps to adjust the reform program to make it more feasible to implement given various constraints, which effectively means using the triggers initially set up in RFs as an "upper bound" for reforms. In cases when DPOs serve as a platform for donor coordination for predictable transfer of donor resources on regular basis, flexibility is used to ensure timely disbursements (i.e., Tanzania and Vietnam PRSCs). In those cases, typically the rigor of the agenda would be adjusted to make DPOs implementable within a certain period of time. In other cases when there is a high premium on timely disbursement (e.g., crisis support), the RFs may undergo changes to ensure such timely disbursement (i.e., Vietnam Public Investment Reform DPOs [PIR]). In some cases flexibility would be required to keep the program going even though the reform momentum is weaker than expected. This may also happen in contexts where series are important for maintaining the dialogue between the Bank and the client.

Box 5.1. An Example of a Partially Met Trigger without Institutional Depth

A prior action—"Ministry of Health has issued a regulation to establish a centralized database as a step toward unifying the national licensing system of healthcare professionals"—was approved as a partial implementation of the trigger on "adopting national standards and establishing a unified licensing system for health care practitioners" in a programmatic series. A project-level evaluation by the Independent Evaluation Group found that these two actions are at different levels of institutional depth and criticality. The prior action was weak, and it should have not been accepted as partial implementation of the trigger.

Source: Case Study of Vietnam PRSC Second Series.

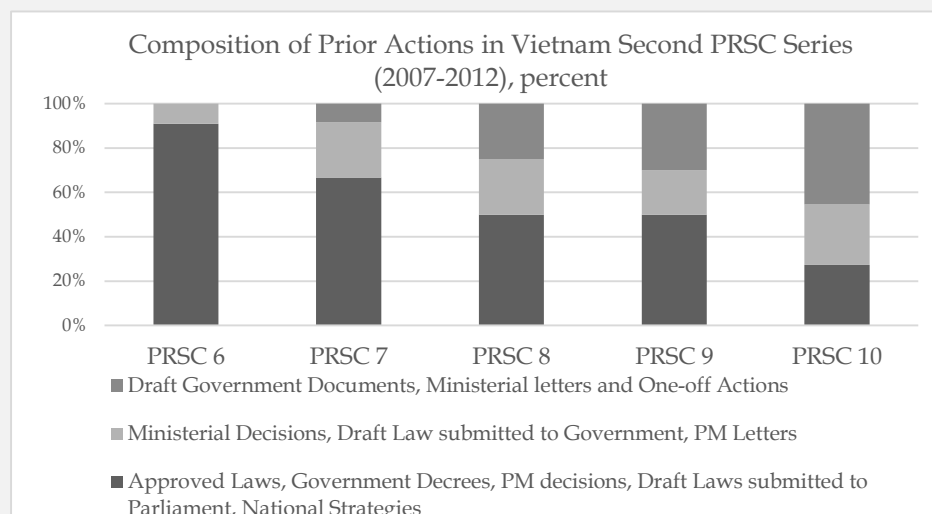
5.6 Flexibility in the programmatic approach can sometimes be excessive. In some programmatic series the quality of RFs suffered substantially due to adjustments in the program (Tanzania PRSC and Vietnam PRSC, PIR). PAs might become weaker, and some key actions may be dropped from the policy matrix, thus undermining the relevance of continued intervention (Box 5.2). The cost of terminating a programmatic series pre-maturely as a result of weaker-than-expected reform momentum can be high for the Bank, as it may undermine relationships and the dialogue with the client. This

5. RESULTS FRAMEWORK IN PROGRAMMATIC APPROACH

may result in weakening the RFs of the series. IEG evidence suggests that this might be an exaggerated concern. If the reform momentum supported by a series is already weak it would be hard to maintain genuine dialogue with the client, assuming that the DPOs addresses the key challenges countries are facing. A DPO series may weaken also in cases when timely disbursements are a priority. Although predictability can be important in certain contexts, in cases where DPO financing is not a significant part of the budget, the rationale for greater flexibility from the Bank's side is less clear.

Box 5.2. Example of Weakening of Design in Programmatic Series

Vietnam PRSC Second Series is an example of gradually weakening quality of PAs. The shift from “policy” to “implementation” and “process” PAs is not a result of sequencing of borrowers' actions. The key reason is that many expected policy actions were dropped and were substituted by actions of less criticality.



Source: Case Study of Vietnam PRSC Second Series.

Special attention must be paid to so-called “partially met” triggers. Often, triggers are only partially met. The Guidance Note requires that changes in triggers should be fully justified and reported. It is up to the Bank to decide if changed triggers are sufficient for the reform program, and this is a context specific process. IEG's evaluative approach in project level evaluations and case studies allow one to assess (also with the benefits of hindsight) if partially met triggers that became PAs can be considered as a step forward toward a program's results. There are good examples of partially met triggers that helped to push the program forward without compromising the program's objectives. In some cases, however, partially met triggers were clearly not sufficient for reform implementation but were accepted as PAs by the Bank (box 5.1). Some partially met triggers that became PAs would have hardly been proposed as PAs at the onset of operations because of insufficient institutional depth.

6. Lessons for DPO Monitoring Frameworks

6.1 The monitoring framework is an integral part of the Results Frameworks (RFs) for development policy operations (DPOs). Because of inherent attribution problems in DPO RFs, monitoring and evaluation (M&E) has been a challenge with many Bank DPOs and with policy lending operations of other International Financial Institutions. On one hand, broad indicators of country outcomes, such as growth and poverty reduction, are controversial because of attribution. On the other hand, choosing indicators that are too output oriented (identical to or similar to the prior actions [PAs]) also would be problematic. Creating a set of relevant indicators in many cases is hampered by weaknesses in country data, while arrangements to generate or collect data specifically for the M&E framework of DPOs may not always work.

6.2 M&E frameworks for DPOs underwent changes in recent years. Substantial efforts have been made to improve M&E frameworks in the Bank. The review of Implementation Completion and Results Report Reviews (ICRRs) and project level evaluations, including the review of a random set of DPOs still under implementation (approved in FY13–14), indicate that there may be an improvement in the M&E framework of DPOs. In particular, the link between PAs and results indicators seems to have strengthened to an extent that, in most DPOs, each results indicator is specifically associated with a prior action. This approach is not directly recommended by Operations Policy and Country Services, but this review finds it a good step forward given the increased importance of PAs in DPO RFs. In addition, among newer DPOs there are fewer indicators with attribution problem than there are among older DPOs.

6.3 Reporting on results indicators appears to be improving as well. Among nine available Implementation Completion and Results Reports (ICRs) for DPOs closed in FY14, none reported missing values on results indicators while five out of 20 ICRs prepared for DPOs closed in FY13, and 10 out of 35 ICRs for DPOs closed in FY12 had missing values. Problems with the reporting of results indicators at completion is less of a concern than in DPOs implemented in the late 2000s and early 2010s (especially Poverty Reduction Support Credits), some of which had as many as 50 percent of results indicators without values at completion. However, as the result of relatively small number of DPOs closed in FY13–14 that were validated by the Independent Evaluation Group (IEG) as of April 2015, it is not clear the extent to which these improvements affected IEG’s rating of the quality of M&E in DPOs¹⁸.

¹⁸ The ratio of DPOs closed in FY07–12 with “high” or “substantial” M&E rating was relatively stable at around 40 percent.

6. LESSONS FOR DPOs MONITORING FRAMEWORK

6.4 The presentation of M&E frameworks was simplified. In recent DPOs, M&E frameworks are fully merged with the programs' policy matrices, with few exceptions. The use of intermediate indicators has been declining in general. An interesting example is the contrast between Malawi 2012 and Malawi 2013. There were intermediate results indicators in the Malawi 2012 stand-alone operation, while the Malawi 2013 operation, a programmatic series, had none. The old approach of presenting some of the indicators as Program Development Objective indicators, effectively creating a "two-tier" M&E framework, has been largely phased out. With few exceptions (e.g., Malawi, Côte d'Ivoire), the programs approved recently (FY13–14) do not make distinctions in M&E. In general, changes toward simplification have contributed the clarity and consistency of the M&E. But the ultimate quality of M&E depends on whether the choice of M&E indicators is strategic.

6.5 Notwithstanding positive changes, weaknesses in the model of presenting the RFs discussed earlier can affect the M&E framework. Presentation of DPO RFs should follow a simple logic that links the policy actions to the outcomes and then to the program objective, while the outcome indicators should measure the extent to which this casual chain materializes. The choice of result indicators for policy and intuitional reforms can be challenging, and some trade-offs will be inevitable, such as outcome-versus output-oriented indicators. A program's M&E will benefit greatly from more strategic choice of indicators that are outcome oriented but at the same time are not too broad to undermine attribution of the results to DPOs. Output-oriented indicators, if there is no practical alternative, may still help to capture the extent of desired change if additional supporting information is provided at completion in ICRs. However, use of such indicators creates problems beyond M&E, in cases when there is no clear statement of outcome in the policy matrix, and the burden of presentation of the result chain shifts to output oriented results indicators, which undermines the quality of RFs.

7. Quality of Reporting at Completion and Implications for Validation and Evaluation

Lessons from ICRs

7.1 Implementation Completion and Results Reports (ICRs) are important elements in the project cycle with great potential to enhance the World Bank's focus on results. ICR preparation is of essential importance for development policy operations (DPOs). The time between approval of the program and completion of the ICR is a critical period where expected changes should materialize. In fact, many important policy actions that follow prior actions (PAs) take place after the formal closing of the program. Policy dialogue in many cases goes beyond program closing.

7.2 The experience of the Independent Evaluation Group (IEG) with validation of ICRs indicates that there are gaps in reporting on DPO results at completion. Although DPOs and their Results Frameworks (RFs) have come to be centered on PAs, the ICRs in general have not followed this pattern. Shortcomings in the ICRs include insufficient information on the nature of the actions (e.g., dates of implementation not provided in program documents) and lack of a comprehensive account of follow-up actions to PAs during implementation that leads to the achievement of outcomes. Unfortunately, many ICRs are missing a detailed storyline on PAs even though they are the key element of the DPOs' conditionality framework. For example, it is typical for ICRs not to mention the status of draft documents prepared as PAs or not to reflect on agency level actions selected as a prior action for DPOs. In many cases, the reason why this information is not presented in the ICR is not because of slow implementation or policy reversal but because of lack of understanding of the value of reporting on PAs and on what happened after these actions.

7.3 Strengthened ICR standards should enhance DPO design. Increased emphasis on reporting on PAs should over time enhance the Bank's corporate standards for identifying PAs, further strengthening the Bank's focus on results through more relevant and critical PAs. It would also likely be beneficial if PAs could be followed up with in the reporting framework of the country assistance strategy to ensure continuity in the monitoring of institutional development in areas supported by DPOs.

7.4 Change in the timing of ICR preparation for DPOs was an important improvement. IEG's experience indicates that six months for preparing ICRs for DPOs was insufficient in many cases for capturing the extent and sustainability of the impact from policy changes supported by DPOs. This was an impediment especially for stand-

alone operations. This short time-frame affected the ability of ICRs to capture cases of policy reversal or reform slowdown and also affected data collection for monitoring results. The Bank's move to a 12-month ICR preparation timeline (since FY15) is effectively an extension of the DPO's life cycle and addresses some of the issues related to the timing of the instrument. With this change, the cycle of stand-alone operations could reach 2.5-3 years, which provides sufficient time to carry out many reform initiatives.

Implications for IEGs Validation and Evaluation Methodology

7.5 Certain aspects of IEG's validation approach should be modified. Better alignment with recent changes in the instrument, such as increased emphasis on PAs and simplification of the monitoring and evaluation (M&E) framework is needed. To avoid possible misinterpretations IEG reviews should take into account that DPOs no longer use "benchmarks" or intermediate indicators.

7.6 Proper presentation of PAs in ICRRs is important. Not all ICRRs mention the PAs. Based on synthesis of all ICRRs of DPOs prepared in 2011–2014, this review finds that the informative and evaluative value of ICRRs that provide detailed description of PAs is higher than those that do not. Because of IEG's approach that ICRRs should be self-explanatory documents, a summary presentation of prior action (preferably in the section on policy areas) would improve the informative value of ICRRs.

7.7 IEG validation should focus more on the analysis of PAs for an outcome rating. Analysis of PAs and their relevance and criticality for the outcomes should be one of the key elements for assessing the relevance of design. With respect to assessment of achievement of objectives, ICRRs should focus on attribution of outcomes to the PAs and the Bank's other inputs, such as policy dialogue. Reconstruction of the chain of events that follow the implementation of PAs leading to the expected results is important for a comprehensive assessment of a program's efficacy. The analytical framework used for this study to assess the quality of PAs that captures their various qualities may be explicitly or implicitly used in the ICRRs.

7.8 The IEG's ICR validation of DPOs will benefit greatly from changes in the corporate approach to RFs' presentation and to the reporting of outcomes at completion. Improving statements of objectives is important for consistency of IEG validation. Additionally, evaluators often have difficulties in separating various aspects of the RFs to be covered under the relevance of design and quality of M&E. ICRR guidelines suggest that the results indicators should be reviewed in M&E while relevance of design should review relevance of stated outcomes. This becomes a

7. LESSONS FOR REPORTING RESULTS AND IMPLICATIONS FOR VALIDATION AND EVALUATION

problematic issue because of weaknesses in the current model that often confuses the outcomes and results indicators. Improving the policy matrix along the lines presented in chapter 2 would improve the process of validation and ensure consistency in the approach. With respect to the quality of ICRs, coordination between Operations Policy and Country Services and IEG is important to ensure sound and comprehensive reporting and validation of DPOs.

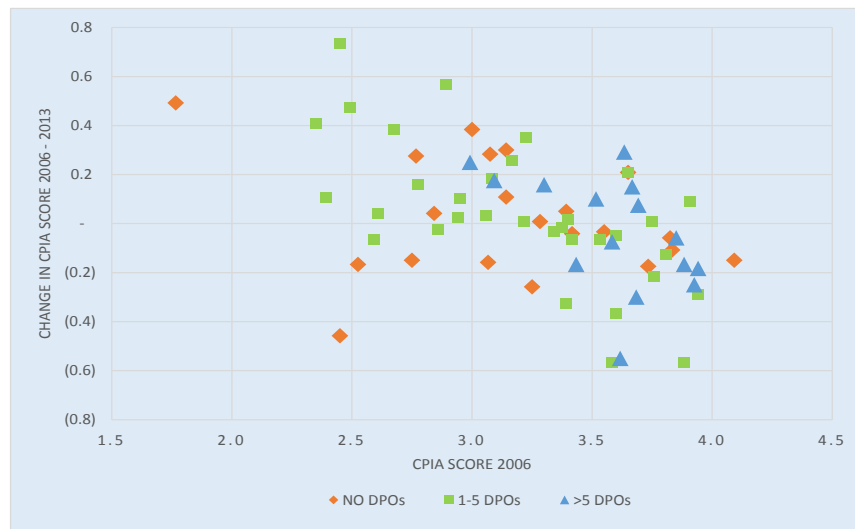
References

- Francisco, Maria. 2012. *2012 Development Policy Lending Retrospective: Results, Risks, and Reforms*. Washington, DC: World Bank.
- Geli, Patricia, Aart Kraay, and Hoveida Nobakht. 2014. Predicting World Bank Project Outcome Ratings. World Bank Policy Research Working Paper 7001.
- Geli, Patricia, Peter Moll, and Pablo Saavedra, 2015. Correlates of Success of Development Policy operations, World Bank Policy Research paper.
- IEG. 2010. Poverty Reduction Support Credits: An Evaluation of World Bank Support. Independent Evaluation Group, The World Bank.
- IEG. 2012. Designing a Results Framework for Achieving Results: A How to Guide.
- IEG. 2010. Poverty Reduction Support Credits: An Evaluation of World Bank Support.
- IEG. 2010. PRSC Results Framework Evaluation
- IEG. 2013. Guidelines for Reviewing World Bank Implementation Completion and Results Reports.
- IEO 2005. Evaluation of Structural Conditionality in IMF-Supported Programs, Issues paper for an Evaluation by the Independent Evaluation Office of the International Monetary Fund.
- OP/BP 8.60. Development Policy Financing (updated July 2014).
- OPCS 2013. Guidance Note for Development Policy Operations: Macroeconomic Policy Frameworks and Collaboration with the IMF in Development Policy Operations. September 2013.
- World Bank. 2011. Good Practice Note for Development Policy Lending: Results in Development Policy Lending. Washington, DC: World Bank
- World Bank. 2011. Good Practice Note for Development Policy Lending: Results in Development Policy Lending. Washington, DC: World Bank.
- World Bank. 2011. *Good Practice Note for Development Policy Lending: Results in Development Policy Lending*. Washington, DC: World Bank
- World Bank. 2011. Good Practice Note for Development Policy Lending: Results in Development Policy Lending. Washington, DC: World Bank
- World Bank. 2011. *Good Practice Note for Development Policy Lending: Results in Development Policy Lending*. Washington, DC: World Bank.
- _____. 2013a. *2012 Development Policy Lending Retrospective Results, Risks, and Reforms*. Washington, DC: World Bank.
- _____. 2013b. *Program Document for a Proposed Development Policy Grant in the Amount of SDR 33.4 Million (US\$50 Million Equivalent) to Republic of Rwanda for a Quality of Decentralized Service Delivery Support Development Policy Operation*. Washington, DC: World Bank.
- _____. 2014. *Implementation Completion and Results Report (IDA-H8530) on a Grant in the Amount of SDR 33.4 Million (US\$50 Million Equivalent) to the Republic of Rwanda for Quality of Decentralized Service Delivery Support*. Washington, DC: World Bank.
- World Bank. 2011. *Good Practice Note for Development Policy Lending: Results in Development Policy Lending*. Washington, DC: World Bank.
- Smets, Lodewijk; Knack, Stephen. 2014. "World Bank Lending and the Quality of Economic Policy." Policy Research Working Paper 6924, World Bank Group, Washington, DC.

Appendix A. Results in Development Policy Operations

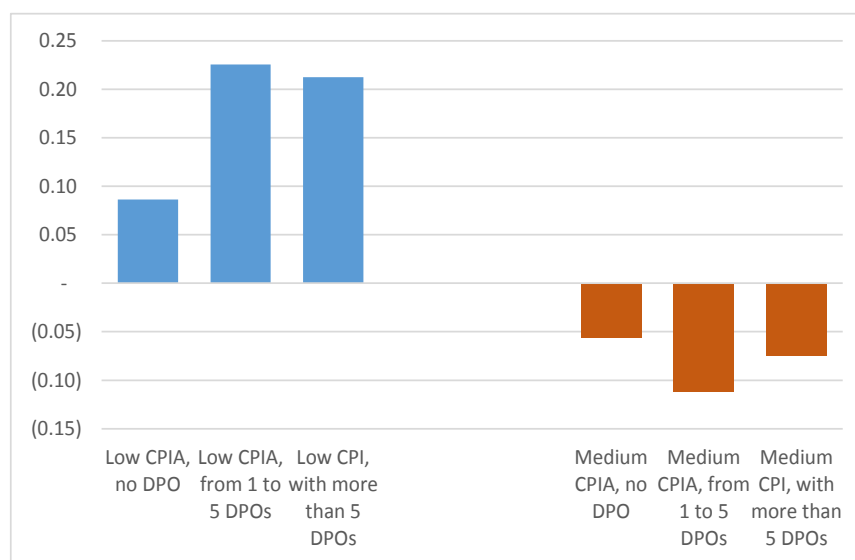
1. There is mixed evidence on the overall effectiveness of the Bank's development policy operations (DPOs) partly because of the difficulties in measuring the impact on country outcomes. By one measure, DPOs appear to perform relatively well. At least three-fourths of all DPOs that closed in FY06–13 were assessed as moderately satisfactory or above by the Independent Evaluation Group in achieving their intended outcomes (outcome rating). The share of DPOs with satisfactory and highly satisfactory outcome ratings has been around one-third of all DPOs in recent years.
2. There is a tradition in the World Bank of linking Country Policy and Institutional Assessment (CPIA) ratings to DPOs for research and evaluation purposes since CPIA ratings are convenient for capturing the quality of institutions required for achieving the Bank's twin goals. In general, the results in policy and institutional change across the International Development Association's eligible countries is mixed¹⁹. Figure A.1 shows that 52 out of 105 countries that had CPIA ratings below 4 in 2006 have experienced a decline in CPIA ratings since 2006. Countries that had relatively high CPIA ranking in 2006 were more likely to regress in subsequent years. There are possible explanations for this including difficulty in achieving more positive changes at relatively high level of institutional development and reform fatigue.

Figure A.1. CPIA change in 2006-2013.



¹⁹ Knack 2014

Figure A.2. CPIA change, by low CPIA, and intensity of DPO



3. There is mixed evidence on the contribution of DPOs to changes in CPIA ratings. Attempts to statistically measure the impact of DPOs on CPIA scores have produced inconclusive results for the subset of countries that are International Development Association recipients. As a simple illustration, figure A.2 shows that better CPIA outcomes are not necessarily associated with DPO operations, judged by both the presence of DPOs and the number of operations. Although in countries with low CPIA ratings (< 3.2) the presence of DPOs is associated with better performance in terms of CPIA change, this association disappears and even reverses in IDA-eligible countries that have aggregate CPIA ratings above 3.2. In other words, the Bank's DPOs do not seem to be associated with positive change in countries with CPIA rating above 3.2, suggesting that the question of DPOs' results orientation and the quality of their Results Framework is still valid²⁰.

²⁰ More analysis is needed to assess the links between Bank's policy lending and the quality of policy institutions.

Appendix B. Example of Policy Matrix

Table B.1. Policy Matrix

Development Objective: Strengthen the government's program to reduce risks resulting from adverse natural events		
Policy Issue	Prior Action	Outcome Indicator
Strengthening disaster risk management institutional and planning capacity	The borrower submitted to its Congress a draft bill 43 under which a national policy and national system for disaster risk management are established, as evidenced by draft law 50 of 2011 published in the borrower's <i>Congress Gazette</i> on August 11, 2011; and the borrower created a national unit for the coordination of said disaster risk management system as evidenced by Decree 4147 of 2011	Increase GoC capacity for disaster risk reduction planning and implementation: <ul style="list-style-type: none"> The new Agency for the Coordination of the National DRM framework has put in place new instruments for interagency coordination. (Baseline: , Target:) Number of formulated municipal plans for disaster risk management under the technical assistance of UNGDR. (Baseline: , Target)
Strengthening the institutional framework and capacity for including disaster risk management in land use planning	The borrower has mandated the identification and zoning of hazard and risk areas and the definition of specific mitigation measures as part of the review process for the issuance of new land use plans and for the plans as evidenced by Article 189 of Decree 019 of 2012revision of existing land use	Improve technical capacities and institutional organization for including risk assessment in land use planning <ul style="list-style-type: none"> Number of municipalities that have reviewed or formulated new plans with the guidelines for disaster risk management in their territorial development plans under the new decree.(Baseline: Target:) Number of regional autonomous corporations assisted in the incorporation of risk in watershed management plans. (Baseline:Target)

Source: Colombia DRM DPO