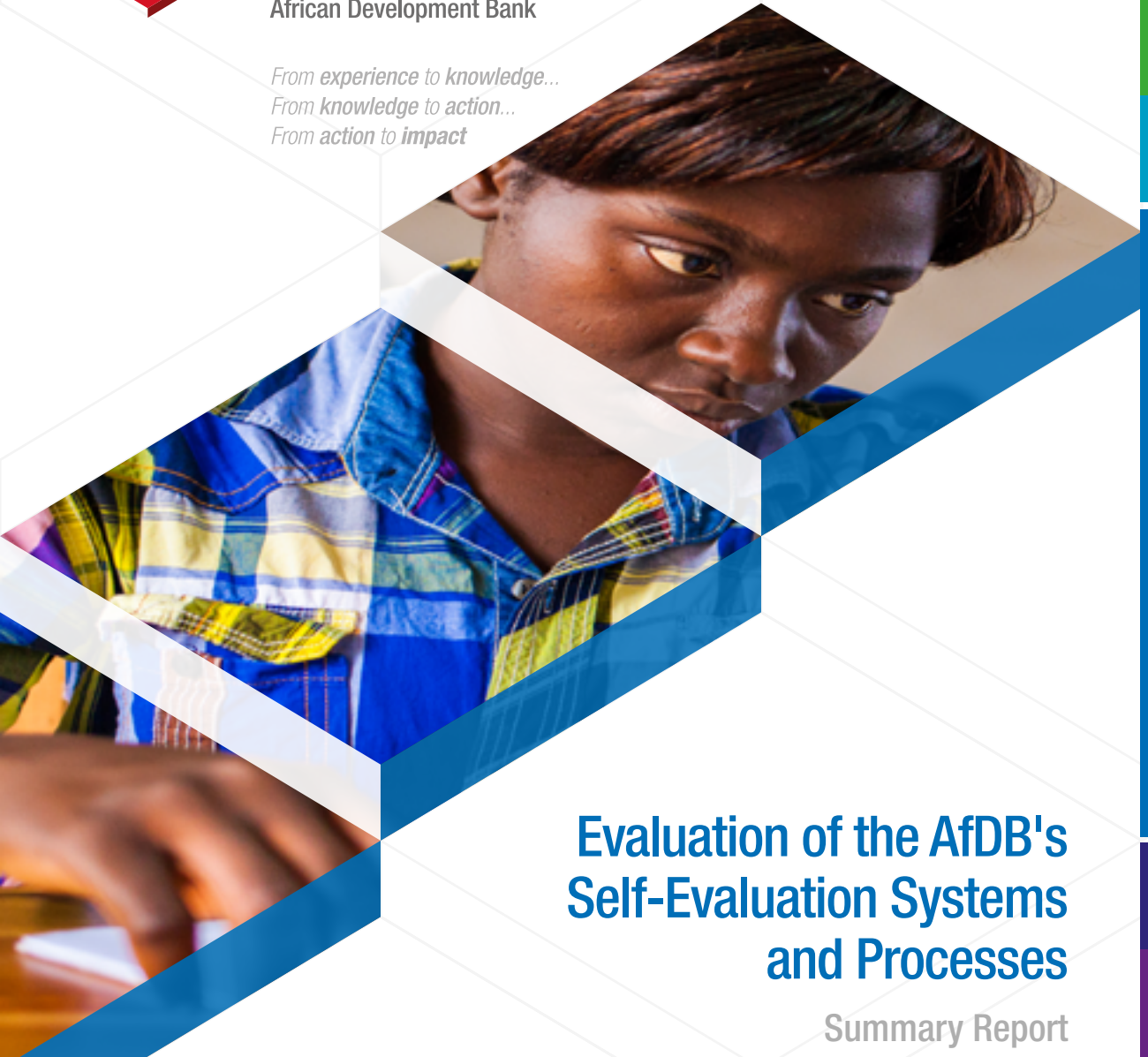




IDEV

Independent Development Evaluation
African Development Bank

From experience to knowledge...
From knowledge to action...
*From action to **impact***



Evaluation of the AfDB's Self-Evaluation Systems and Processes

Summary Report

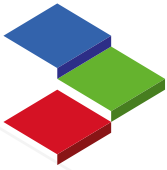


AFRICAN DEVELOPMENT BANK GROUP

April 2020

IDEV conducts different types of evaluations to achieve its strategic objectives

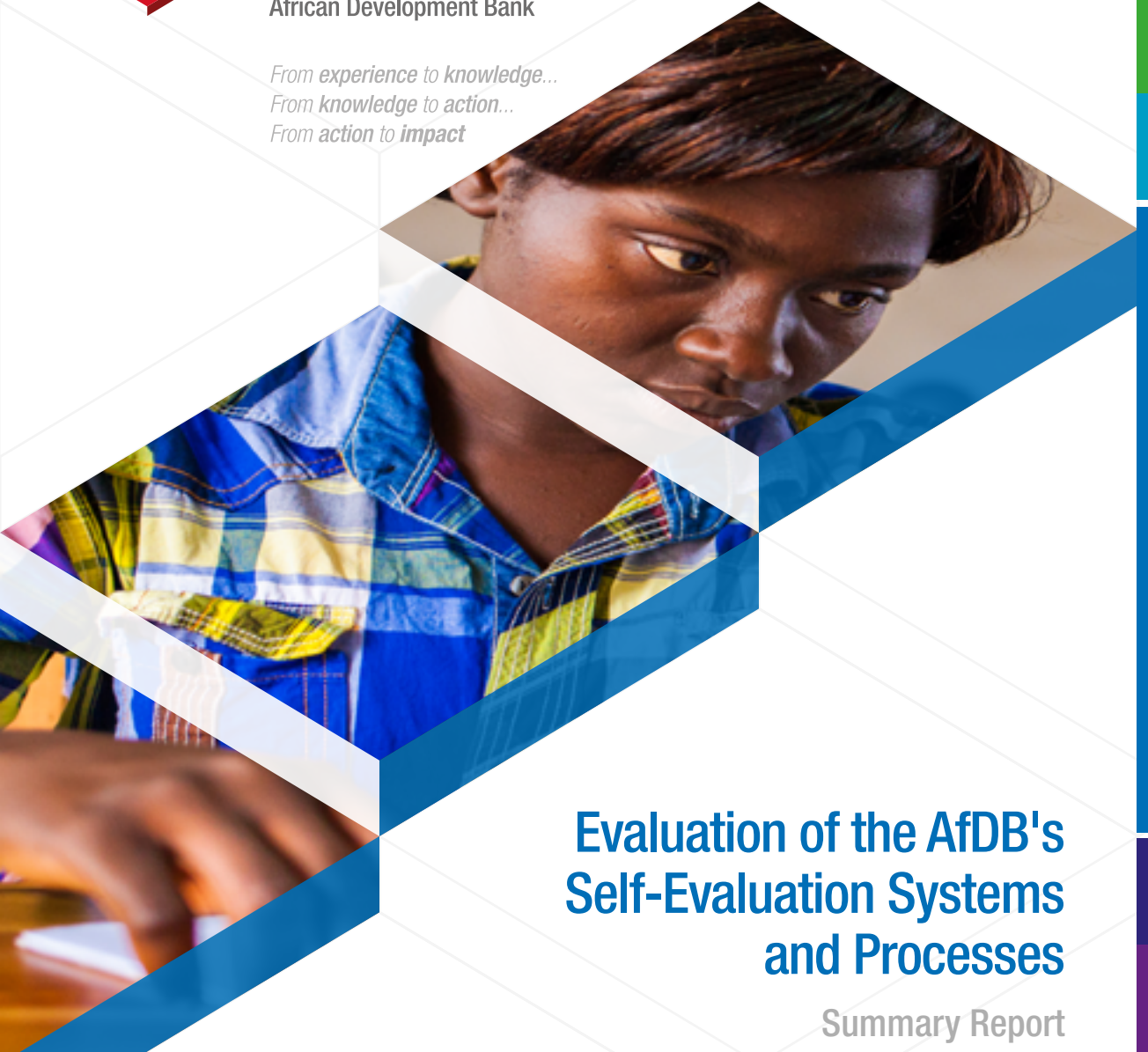




IDEV

Independent Development Evaluation
African Development Bank

From experience to knowledge...
From knowledge to action...
*From action to **impact***



Evaluation of the AfDB's Self-Evaluation Systems and Processes

Summary Report



AFRICAN DEVELOPMENT BANK GROUP

April 2020

ACKNOWLEDGMENTS	
Task manager	Foday Turay , Chief Evaluation Officer, IDEV.1
Team members	Stephanie Yoboué , Junior Consultant, IDEV.1 and Henda Ayari , Team Assistant, IDEV.1
Consultant	Centennial Group International: Team led by Adolfo Brizzi and Hedi Manai
Internal peer reviewers	Andrew Anguko , Chief Quality and Methods Advisor, IDEV.0; Monica Lomena , Principal Evaluation Officer, IDEV.2; Akua Arthur-Kissi , Senior Evaluation Officer, IDEV.2 and Latefa Camara , Evaluation Officer, IDEV.1
External peer reviewer	Marie Gaarder , Executive Director, International Initiative for Impact Evaluation (3ie)
Internal Bank reference group	Claudia Ngassa , Principal Program Coordinator, PINS.0; Carlos Mollinedo , Energy Economist/Policy Specialist, PESR.1; Grace Ochola , Operations Assistant, AHFR.0; Mamadou Tangara , Chief Regional Program Officer, RDGC; Alain Pierre Mbonampeka , Chief Regional Program Officer, RDGE.4; Mukaila A. Ojelade , Consultant, RDGE.4; Noel Kulemeka , Manager, Implementation support Division, RDGS; Adeleke Salami , Principal Macroeconomist, ECMR; Jean Marie Dabire , Principal Country Economist, CODD; Sandrine Alissoutin , Principal Quality Assurance Officer, SNOQ.2; Thierry Kangoye , Principal Strategy and Policy Officer, SNSP.1; Patrice Horugavye , Environmental and Social Coordinator, RDGC.4; Angela Nalikka , Division Manager, PESD.1; Philip Boahen , Chief Agriculture Policy Economist, AHAI.1; Francis Ofoe Konu , Monitoring and Evaluation Specialist, AHWS.0; Tom Mboya Owiyo , Advisor, AHVP; Cyril Blet , Chief Accountability Officer, SNDR.1; Charles Mulingi , Senior Results Specialist, SDNR.1; Samba Kamara , Chief Regional Program Officer, RDGW.0; Evelynne Change , Chief Governance Officer, RDGS.4; Rosemond Offei-Awuku , Chief Development Economist, AHHD.0; Mouhamed Gueye , Chief Development Economist, RDGN.2
Knowledge management officers	Jayne Musumba , Principal Knowledge Management Officer, IDEV.3; Abdoul-Aziz Ouattara , Evaluation Knowledge Assistant, IDEV.3; and Marc Ghislain Bappa , Knowledge Management and Communications Junior Consultant, IDEV.3
Other assistance provided by	Myrtha Diop , Senior Administrative Assistant, IDEV.0; Ruby Adzobu-Agyare , Secretary, IDEV.0 and Helmi Hmaid , Principal Quality Assurance Officer, SNOQ.2
Division manager	Rufael Fassil ; Foday Turay (OIC)
Evaluator General	Roland Michelitsch ; Karen Rot-Münstermann (Acting)

© 2020 African Development Bank Group
All rights reserved – Published April 2020

Evaluation of the AfDB's Self-Evaluation Systems and Processes
An IDEV Corporate Evaluation, April 2020

Disclaimer

Unless expressly stated otherwise, the findings, interpretations and conclusions expressed in this publication are those of the various authors of the publication and are not necessarily those of the Management of the African Development Bank (the "Bank") and the African Development Fund (the "Fund"), Boards of Directors, Boards of Governors or the countries they represent.

Use of this publication is at the reader's sole risk. The content of this publication is provided without warranty of any kind, either express or implied, including without limitation warranties of merchantability, fitness for a particular purpose, and non-infringement of third-party rights. The Bank specifically does not make any warranties or representations as to the accuracy, completeness, reliability or current validity of any information contained in the publication. Under no circumstances including, but not limited to, negligence, shall the Bank be liable for any loss, damage, liability or expense incurred or suffered which is claimed to result directly or indirectly from use of this publication or reliance on its content.

This publication may contain advice, opinions, and statements of various information and content providers. The Bank does not represent or endorse the accuracy, completeness, reliability or current validity of any advice, opinion, statement or other information provided by any information or content provider or other person or entity. Reliance upon any such opinion, advice, statement, or other information shall also be at the reader's own risk.

About the AfDB

The overarching objective of the African Development Bank Group is to spur sustainable economic development and social progress in its regional member countries (RMCs), thus contributing to poverty reduction. The Bank Group achieves this objective by mobilizing and allocating resources for investment in RMCs and providing policy advice and technical assistance to support development efforts.

About Independent Development Evaluation (IDEV)

The mission of Independent Development Evaluation at the AfDB is to enhance the development effectiveness of the institution in its regional member countries through independent and instrumental evaluations and partnerships for sharing knowledge.

Independent Development Evaluation (IDEV)

African Development Bank Group
AfDB Headquarters
Avenue Joseph Anoma, 01 BP 1387, Abidjan 01, Côte d'Ivoire
Phone: +225 20 26 28 41
E-mail: idevhelpdesk@afdb.org
idev.afdb.org

Photography: IDEV, AfDB
Original language: English – Translation: AfDB Language Services Department
Design & layout: A Parté Design

Contents

Abbreviations and Acronyms	v
Executive Summary	1
Management Response	11
Background and Context	23
Purpose, Objectives and Scope	25
Purpose	25
Objectives	25
Scope	26
Methodology	29
Theory of Change	29
Evaluation Questions	31
Evaluation Methods	31
The Underlying Factors of Self-Evaluation	35
The self-evaluation rationale	35
Compliance	35
Candor	38
Managing Ratings	40
Enforcement and Incentives to Ensure Compliance	44
Main SESP Characteristics and Outcomes	49
Performance Management	49
Promoting Accountability	54
Learning	60
SESP adequacy to specific cases	62
Conclusions and Recommendations	65
Conclusions	65
Recommendations	66
Annexes	69

Contents

List of figures

Figure 1:	Self-evaluation system and processes - Theory of Change	30
Figure 2:	SESP incentive structure	47

List of tables

Table 1:	SESP Reporting tools and performance	8
Table 2:	Summary evaluation matrix	31
Table 3:	Projects with satisfactory DOs and efficiency indicators (3 and above)	39
Table 4:	Satisfactory cut-off rating and disconnect	41
Table 5:	Number of projects within rating ranges (out of 137 projects)	42
Table 6:	Bank and Borrower performance (out of 137 projects)	44
Table 7:	NSOs evaluation dimensions	45
Table 8:	Quality of PCRs	56

List of boxes

Box 1:	Definition and overview	27
Box 2:	QoS Evaluation findings on compliance (public sector projects)	37
Box 3:	Candor: Highlights from the QoS evaluation	39
Box 4:	Key issues observed in reviewing the sample of five NSOs	40
Box 5:	The unintended consequences of the rating system	42
Box 6:	QoS - M&E - Desk Review	50
Box 7:	QoS - Follow-up action - Desk review	53
Box 8:	Accountability - QoS Survey Results (83 public projects) - Desk review	55

Abbreviations and Acronyms

ADER	Annual Development Effectiveness Report	OM	Operational Manual
ADF	African Development Fund	PAR	Project Appraisal Report
ADOA	Additionality and Development Outcomes Assessment	PBO	Program-based Operation
AfDB	African Development Bank Group	PCR	Project Completion Report
AsDB	Asian Development Bank	PCREN	Project Completion Report Evaluation Note
ASR	Annual Supervision Report	PINS	NSO and Private Sector Support Department
BTOR	Back-to-Office Report	PIU	Project Implementation Unit
CEDR	Comprehensive Evaluation of Development Results	PPP	Public Private Partnership
CPO	Country Program Officer	PRA	Project Results Assessment
CPPR	Country Portfolio Performance Report	PSR	Project Status Report
CSP	Country Strategy Paper	QA	Quality Assurance
CSPE	Country Strategy and Program Evaluation	QaE	Quality at Entry
DAM	Delegation of Authority Matrix	QoS	Quality of Supervision and Exit
DBDM	Development and Business Delivery Model	RDVP	Regional Development, Integration and Business Delivery Vice-Presidency
DO	Development Outcome/Objective	RISP	Regional Integration Strategy Paper
ECG	Evaluation Cooperation Group	RISPE	Regional Integration Strategy and Program Evaluation
E&S	Environmental and Social	RMC	Regional Member Country
ESAP	Environmental and Social Assessment Procedures	RMF	Results Measurement Framework
ESS	Environmental and Social Safeguards	RRS	Results Reporting System
EVRD	Evaluation Results Database	SAP	System, Applications and Products
IDEV	Independent Development Evaluation	SES	Self-Evaluation System
IED	Independent Evaluation Department	SESP	Self-Evaluation Systems and Processes
IFAD	International Fund for Agriculture Development	SNOQ	Operations Committee Secretariat and Quality Assurance Department
IFC	International Finance Corporation	SOU	Special Operations Unit
IPR	Implementation Progress Report	TA	Technical Assistance
KM	Knowledge Management	TM	Task Manager
KPIs	Key Performance Indicators	ToC	Theory of Change
LoC	Line of Credit	TYS	Ten-Year Strategy
M&E	Monitoring and Evaluation	UA	Unit of Account
MDB	Multilateral Development Bank	WBG	World Bank Group
MTR	Mid-term Review	XSR	Extended Supervision Report
NSO	Non-sovereign Operation	XSREN	Extended Supervision Report Note



Executive Summary

Introduction

The evaluation of the African Development Bank's (AfDB) Self-evaluation Systems and Processes (SESP) comes at a critical time, as the Bank has been going through profound changes, and revisiting its strategic directions and operational processes. Such an effort comes from the realization that, ultimately, it is the quality of operations that determines the capacity of the Bank to achieve development results. As a consequence, the Bank has initiated over the past couple of years an in-depth diagnostic of its Quality Assurance (QA) processes. This diagnostic was complemented by a series of Independent Development Evaluation (IDEV) evaluations that aims to examine the relevance, efficiency and effectiveness of the Bank's QA processes across the project cycle as a way of enhancing their contribution towards performance management, accountability and learning.

This evaluation is conceived with this logic in mind and builds on previous IDEV evaluative work dealing with: (i) quality at entry (QaE); and (ii) quality of supervision and at exit (QoS), as well as a "chapeau"¹ paper coalescing the different findings including the compliance with the Bank's Environmental and Social (E&S) safeguards. Based on existing evidence, Management prepared a QA Implementation Action Plan ("The Plan"), covering five areas of reform² and setting the basis for the improvement of the SESP, building on this evaluation and Management's own diagnosis. This will complete the assessment of the QA framework and will potentially set the stage for institutionalizing "best practices" in a consistent way and positioning the AfDB on a par with comparator agencies.

The SESP are commonly defined as the assessment made of a project, country/regional program, and policy/thematic review by the entity engaged in the activity. As illustrated in the theory of change

(Figure 1), the main SESP tools being used are the implementation progress reports, the mid-term reviews, and the completion reports. The SESP are meant to pursue three main outcomes: (i) performance; (ii) accountability; and (iii) learning. The functioning of the SESP is assessed along three main dimensions: relevance, effectiveness and efficiency. The performance of the main SESP tools is summarized in Table 1 below. While not being SESP instruments as such, corporate reporting through dashboards and the Results Measurement Framework (RMF) are partly fed by the SESP.

While the SESP are carried out by staff/Management, IDEV complements the process through the validation of certain products, such as the completion reports or through independent evaluations of projects, Country Strategy Papers (CSPs) and Regional Integration Strategy Papers (RISPs).

While project performance relies to a large extent on the performance of country governments, partners and local stakeholders, as well as external factors, the SESP are entirely under the Bank's control, as well as in its capacity to adopt and implement recommendations. Ultimately, the evaluation tries to answer the following question: "Do the SESP support performance management, accountability, and learning at the Bank?"; and two underlying sub-questions: (i) how well are the SESP performing?; and (ii) to what extent are the SESP impacting the quality of development results?

The evaluation covers the period 2013 - 2018, and the implementation of the SESP over the time span that goes from project approval to closing (exit). It covers both public sector and Non-sovereign Operations (NSOs). While specific references are made with respect to NSOs in each chapter, the report is mostly focused on public sector operations. In addition to the project- or operation-level analysis, case studies for a sample of CSPs and RISPs have also been carried out.

Main Findings

The Bank's SESP have many positive features. They lay out strong standards and procedures underpinning their functioning, as well as a cogent articulation with the independent evaluation function carried out by IDEV. Annex 5 presents a detailed description of the SESP of comparator institutions that shows a good level of alignment, which was to be expected from members of the Evaluation Coordination Group (ECG)³. Many of the tools and processes in place are consistent with, and as good as, those of comparator institutions. The articulation between self-evaluation, validation and independent evaluations is similar, and the system produces corporate results data that are used to report to the Board. However, this evaluation finds that progress towards a culture of development effectiveness has been mixed, and the potential of the SESP to make an impact on the three fronts of performance management, accountability and learning, remains unmet.

The main weaknesses of the SESP are in the application of the established procedures, standards and norms. While the issues identified as part of this evaluation mirror similar constraints faced by comparator organizations, three factors seem to be affecting the proper functioning of the SESP to a higher degree in the AfDB: (i) low compliance with established procedures; (ii) limited resources for M&E during supervision; and (iii) a deficient in candor and a positive bias in assessing performance. This has affected the credibility of the SESP and contributes to a perception that the system is adding little value.

The AfDB has an independent evaluation policy that was approved in 2016 and amended in 2019. However, the Bank does not have an integrated evaluation policy that covers both the independent function and self-evaluation of the Bank itself (such as in the case of the World Bank Group [WBG] or IFAD). IDEV promotes the use of evaluation findings on specific topics in line with the

demand coming from the Board and Management, which are also the primary users. Evaluation adds value only when its findings are used. The AfDB's independent evaluation policy necessarily focuses on the supply side of evaluation and not the use side. In the absence of a harmonized framework, alignment of strategic approaches, methodologies and processes will remain at risk. Furthermore, IDEV is not the only generator of evaluation findings. All completed operations and country/regional strategy papers (CSPs/RISPs) are self-evaluated by staff.

Effectiveness of the SESP is constrained by its ratings methodology and structure, and the way it is applied to Project Completion Reports (PCRs). Several elements can be highlighted:

- The Bank uses a 4-point rating system⁴, but several discussions have taken place recently on the merits of adopting a 6-point system similar to other comparator organizations (IDB, IFAD, and the WBG). A 6-point system allows for more flexibility and realism in assessing the performance of the large pool of projects falling somewhat in between Satisfactory and Unsatisfactory, as the distance between the two is often perceived to be too wide and too stark, according to staff interviews. However, it was also noted that the 6-point scale does not permit making a clear judgment regarding project performance, and that such a scale would tend to classify performance of most interventions in the two middle categories anyway, which ultimately will be aggregated together with the Satisfactory or the Unsatisfactory categories as being "above or below the bar". Other suggestions made by the Evaluation Reference Group (ERG) or by the 2016 PCR Synthesis Report are in favor of an odd-number points system (3 or 5), as this will allow to better account for the middle space of the distribution (assuming a symmetrical profile). However, an odd number rating scale would be against the agreed Good Practice Standards of the ECG⁵.

- The main risk associated with any rating scale is whether it might lead to more positive ratings without supporting evidence and subsequently to an increased “disconnect” with IDEV ratings. This happens to be the case in the Bank. Any rating system will have pitfalls and, whether a 3, 4, 5 or a 6-point rating is adopted, it may not lead to significant improvements unless other measures are put in place to improve the reliability of the system and the implementation of more rigorous procedures for the generation of the required evidence.
- The use of simple averages of sub-ratings and dimensions (relevance, effectiveness, efficiency, sustainability) can somewhat distort overall ratings. Some institutions (e.g., IDB) put a higher weight on effectiveness, or do not allow a positive rating with lacking relevance or effectiveness. Defining the cut-offs for a high rating is also important. The AfDB defines “satisfactory” (or better) overall performance as a rating above 2.5, meaning that half of the dimensions being above the line and the other below the line is sufficient. Other institutions use a higher bar (e.g., AsDB requires an average of 2.75, i.e., 3 of 4 dimensions have to be above the line), or they do not use averaging, hence they avoid having to deal with decimal numbers (WB, IFC, IFAD, IDB).
- There is a significant difference on project performance depending on how “satisfactory” is defined. At 2.5 and above, 97 percent of the 137 PCRs analyzed⁶ are considered to be satisfactory, while at 3 and above it is 80 percent. Similarly, the “disconnect”⁷ between the overall rating assigned by staff in the PCRs, and the lower rating provided by IDEV validation in the respective Project Completion Report Evaluation Note (PCREN) is 15 percent at 2.5 and above and 39 percent at 3 and above. This raises questions regarding the credibility of the PCR ratings.
- In recent years, the AfDB has reported in its Annual Development Effectiveness Review (ADER) the PCR/Extended Supervision Report (XSR) ratings instead of the PCREN/Extended Supervision Report Note (XSREN) ratings provided by IDEV. This is mainly because of the untimely availability of the PCREN/XSREN ratings⁸. The use of the PCR/XSR ratings especially for RMF indicators on portfolio performance departs from the practice of comparator institutions.
- Bank and Borrower performance ratings in the PCR are very different from IDEV ratings in the PCREN. It has proven difficult for staff to rate the Bank’s performance negatively. This is seen as closely associated with their own performance and that of their clients, with whom staff try to maintain a good relationship and naturally they do not want to be seen as finger-pointing. The large perceived step between a Satisfactory and an Unsatisfactory rating has likely also contributed to this gap.
- The downgrading of ratings by IDEV is a source of tension and is often not accepted by staff, who argue that validation is a desk-based exercise that does not account for field realities. Ratings will remain a controversial subject as long as they are viewed as a tool to pass judgment on staff performance. Consultations between staff and the IDEV team around the PCR validation process are limited to the provision of feedback on draft PCR validation. In the opinion of almost all the staff interviewed, the rating methodology needs revision. Whether the system is too heavy on ratings or not remains to be seen, but at a minimum there should be an attempt to mitigate the perception that ratings are easy to “game”, that IDEV’s validation is out of context and mostly biased in nature, that the disconnect with IDEV can be largely ignored, and that the rating methodology is too rigid and bureaucratic.

As illustrated in the theory of change (Section III), the underlying logic of the SESP is that it can play an important role in improving performance management, accountability and learning. The rest of the findings are organized along these three following main outcomes.

Performance Management

The lack of candor in self-evaluation, particularly in Implementation Progress Reports (IPRs)/Annual Supervision Reports (ASRs) and PCRs/XSRs, can be explained among other things by the lack of a proper incentive structure.

This is corroborated by various sources, including the Operations Committee Secretariat and Quality Assurance Department (SNOQ) Quality Retrospective Report, the QoS evaluation, the 2016 and 2017 PCR Synthesis Reports, staff interviews and the case studies. The perception that project performance is equated to staff performance undermines the motivation to rate poorly-performing projects candidly. The effectiveness of the SESP is undermined by their being viewed as a compliance mechanism that is driven by box-ticking, protecting one's reputation, and relying on weak generation of evidence (M&E and results frameworks). Candor is also undermined by weak accountability mechanisms. Finally, there is little recognition that being identified as a "problem fixer" could motivate staff towards greater proactivity in raising issues and corrective action.

A number of issues constrain the contribution of the SESP to improving portfolio performance.

- The Bank's culture, incentives and institutional Key Performance Indicators (KPIs) are skewed in favor of lending approvals, similar to other comparator institutions, with limited emphasis on quality and development results. This issue has been recognized by Management and is being addressed through the QA Action Plan;
- There is insufficient attention to incentives that support a culture of quality and results. Opportunities for recognizing, celebrating, internalizing and learning from good quality M&E, results frameworks, proactivity in addressing issues or project restructuring, and successful implementation remain limited (this is also being addressed by the QA Action Plan);
- There are weaknesses in M&E systems and how they are articulated with baselines and results frameworks. Many PCR ratings were downgraded by IDEV validations for lack of evidence that would support a particular assessment;
- Excessive focus on accelerating project approvals leads to critical design activities being rushed or carried over into implementation (procurement plans, feasibility studies, validation of E&S assessment studies, the setting-up of the Project Implementation Unit [PIU] baselines). This results in early implementation delays, which require a stronger SESP to fix issues from the start;
- There is a strong tendency to avoid addressing issues through formal project restructuring because the transaction costs are considered to be too high. This results in a failure to introduce corrective measures and leads to the retention of appraisal targets that are no longer in line with the project reality;
- Increased decentralization and the move towards continuous implementation support, together with the establishment of a new position of Implementation Support Manager for public sector projects, are positive developments. However, new roles and the division of responsibilities between staff at HQ and at the country/regional level need clarification;
- The IPR ratings of the current portfolio feed into the Country Portfolio Performance Reports (CPPRs), which in turn also affect the assessment of the CSPs, since they are presented in tandem. The Development Objectives (DO) and Implementation Progress (IP) ratings of the last IPR also migrate to the PCRs. Therefore, the compliance and candor issues affecting the IPRs

permeate through other SESP outputs and all the way to the Results Measurement Framework (RMF). In addition, issues of consistency between CSPs and CPPRs' assessment methodologies have emerged recently that require revision (currently under way);

- In the case of NSOs, the lack of clear measuring tools of the progress towards development objectives and the absence of a rigorous and institutionalized M&E system lead to poor tracking and reporting of achieved results, and reduce the likelihood of effective risk mitigation during implementation. This makes it difficult to assess, analyze and report adequately on portfolio performance. These issues are now being contemplated under the NSO QA Action Plan (September 2019) adopted by Management.

Project teams and task managers (TMs) are fundamentally motivated to help clients deliver results and, by and large, are committed to the supervision task. However, most TMs and investment officers are overstretched, and the additional support required to address issues and help in the proper implementation of the SESP has been wanting. The Bank seems to be short of staffing and skills required to implement the SESP efficiently and effectively. This is particularly the case for M&E and E&S safeguards during supervision, as also documented in the recent evaluation report on the integrated safeguards system (ISS)⁹.

Accountability

Low compliance stems from insufficient accountability mechanisms and deficient visibility. In the absence of reliable information, the SESP lose their credibility. The current enforcement and incentive systems fail to prevent staff from ignoring the rules that suit them least and getting away with not generating the required outputs. Moreover, the Bank has been lagging behind in the development of an IT-supported online portal that can provide access to portfolio and SESP data and raise the visibility and the efficiency of the system.

Providing real-time information and compliance data to line managers will enhance accountability for supervision, and completion of IPRs/ASRs and Mid-term Reviews (MTRs), while easing the burden on TMs by simplifying and streamlining reporting requirements. It will also facilitate more effective portfolio reviews and planning exercises. The roll-out of the Results Reporting System (RRS), as part of the Integrated QA Action Plan for both the public sector projects and the NSOs, is meant to address this issue. Once completed, it is expected to facilitate automatizing the reporting of results, the timely escalation of issues to senior management, accountability, reliability of information, efficiency of reporting at the corporate level, and transparency at implementation.

Reporting tools such as Dashboards and the RMF have proven to be powerful accountability mechanisms and should be enhanced to cover indicators of compliance. The fact that the disconnect between Management's ratings and IDEV's ratings is not reported takes away an important element of accountability. In addition, very little information is conveyed to the Board in the RMF with respect to the performance of NSOs. While there is a need to take into account issues of confidentiality, this could be addressed through aggregate reporting.

Accountability processes have placed insufficient focus on the quality of monitoring and closure. The IPR/ASR and the PCR/XSR are not fully considered a decision-making or a learning tool, but rather an administrative requirement with inadequate accountability. Since IDEV does not validate IPRs/ASRs, there are few opportunities for contestability of ratings, as evidenced by the fact that IPRs are not much discussed nor systematically reviewed by Management. Line managers are not systematically held accountable for quality checks at supervision. The role played by the Portfolio Management Division of the NSO and Private Sector Support Department (PINS) is akin to a dual accountability approach and provides some degree of arms-length review and contestability.

For public sector projects, the recent addition of the Implementation Support Manager position in each region could help engage on portfolio issues, liaise with the sector divisions, and provide the regional and country perspective, including country program officer (CPO) involvement for better convergence with country portfolio management.

A consistent and harmonized framework that allows for the assessment of performance throughout the project life from origination to independent evaluation is lacking. This would allow linking the SESP and its tools with the front-end portion of the QA process (quality at entry) and ensure that the same approach and indicators are being used throughout the process, in progress reports and in PCRs and XSRs. This will also make clearer where the lack of candor comes in and allow for “no surprises” by providing predictability on the assessment metrics being used. This also means keeping results frameworks up-to-date in case of changes to the project environment and, on the NSOs side, including monitoring indicators in line with those adopted by the Additionality and Development Outcomes Assessment (ADOA) framework.

Templates and formats for a number of SESP tools are not sufficiently differentiated and adapted to specific circumstances. There is room to make the templates more efficient and user-friendly and avoid redundancies especially between IPRs/ASRs and Back-to-Office Reports (BTORs), as also noted by the QoS evaluation. IPRs are not adapted to special investment vehicles such as Program-based Operations (PBOs) and technical assistance (TA) projects, nor to fragility situations. The PCR and PCREN templates are overly repetitious, too long and some sections duplicative, as also evidenced by the Final Synthesis Reports of the 2016 and 2017 PCR validation (2019). Templates are not designed for optimum management attention and do not focus on priority issues or priority actions needed.

The capacity of the SESP products to report and address specific issues is weak in the areas of safeguards, gender, climate and

fragility. In particular, as reported by the recent ISS evaluation, the Bank's supervision reports and BTORs do not capture the key E&S information to allow compliance checks. E&S information found in the available supervision reports is scanty except when a potential high corporate risk materializes. Other studies conducted by SNSC corroborate the point that, once a project is approved, the Bank's internal reporting system offers very little information regarding the E&S follow-up. The gender dimension of M&E systems is particularly weak, with average ratings in the PCRENS below 2. Project teams do not have adequate support in these areas or coaching on SESP requirements to mitigate the effect of the high staff turnover that the Bank has experienced in the past couple of years. The newly launched Operations Academy should help address this issue once all modules are developed.

Completion reports of CSPs/CPFRs and for RISPs are usually not validated by IDEV before being submitted to the Board, unlike comparator institutions or similarly to PCRs and XSRs, which are validated and submitted to the Board in the form of a synthesis paper. IDEV has carried out validation of one CSPs and one RISP on a pilot basis and is assessing whether there is scope for expanding the process. Management is also in the process of reviewing the CPPR methodology and guidelines that should address this point, including the differentiation between CPPRs and CSPs¹⁰.

Learning

Despite the fact that performance management and accountability aspects of the SESP are weak, they have overshadowed or even undermined learning objectives. This finding resonates with the situation in comparator institutions. There is a fine line between the search for accountability and learning. Strategy papers (country or thematic) are more conducive to learning, possibly because no ratings are involved. Disagreement over ratings between staff and IDEV further undermines the incentive to promote learning. If the PCR/XSR is seen as a tool to judge the TM, it will undermine candor

in ratings and learning opportunities. The rating itself could become an obstacle to learning because it potentially makes the discussion unnecessarily contentious and personalized.

There is no single place where SESP products and information can be accessed. The SESP has not benefited from the availability of a solid repository of knowledge that is mined and shared regularly by staff. This is in the making with the advent of the RRS, which will include data from all SESP products. Incentives, combined with new forms of learning and templates, may be needed, providing solutions and lessons irrespective of how it may be self- or independently assessed in a single project context.

The quality of PCRs, as measured by IDEV's validation process, is variable, but average ratings are low (around 2.8), which hinders learning opportunities. While over three-quarters (77 percent) of PCRs are good when measured against the current 2.5 rating threshold (2016 and 2017 validations), only about half (52 percent) are good when measured against a 3-rating threshold, i.e., "fully" satisfactory. The majority of PCRs are prepared by consultants. Most of the TMs interviewed think that they should not do their own PCR for reasons of conflict of interest. However, while using consultants under the supervision of TMs may provide some level of arms-length review, candor remains an issue and consultants do not follow the same standards, raising issues of comparability and quality. PCRs are typically given lower priority by staff and there is little vetting on the choice of consultants.

There is no systematic feedback from the SESP of NSOs that provides success or failure stories based on an assessment of achievements in reaching development outcomes and profitability. Lessons learned are not institutionalized and documented to ensure a strong capitalization of lessons learned. However, IDEV's high level evaluations are contributing to capturing lessons and generating knowledge that enable new strategic orientations for NSOs and private sector development.

The SESP on its own cannot cater for the array of learning needs of the Bank. Learning through the SESP is not sufficiently complemented by other sources of knowledge, such as impact evaluations, thematic reviews, etc. There is a need for more creativity in terms of dissemination and sharing lessons with other countries, in the same sector or on similar specific issues. Periodic workshops or events, as recommended in the PCREN guidelines, organized with relevant staff to enhance feedback on findings drawn from PCR/XSREN, rarely occur. If SESP documents are not sufficiently evidence-based and events are not seen as a safe space where people are willing to learn from success and from failure, incentives to learn will remain insufficient.

Recommendations

From the findings above, and considering the Bank's Integrated QA Plan, Management should:

Recommendation 1: Review the AfDB's results measurement framework and evaluation frameworks across the project cycle to ensure (i) internal consistency throughout the AfDB's results measurement and reporting system from operation to corporate level, and (ii) that there is harmonization between sovereign (SO) and non-sovereign operations (NSO).

Recommendation 2: Review and revise, in close collaboration with IDEV, the PCR/XSR ratings methodology in use including the ratings scale and guidelines in order to improve the reliability of the ratings system for all operations, and to better align SO and NSO.

Recommendation 3: Enhance the accountability for SESP products by developing appropriate indicators and explicitly incorporating these indicators in the AfDB's top-level corporate KPIs, in VPS and Directors performance agreements, and in staff performance discussions. The focus should be on delivery, timeliness, quality, usefulness and proactivity; encouraging candor in flagging and addressing emerging issues.

Recommendation 4: In the context of the AfDB's rightsizing, **review the SO and NSO staffing levels for M&E, and quality assurance activities** in order to ensure that the self-evaluation function is appropriately staffed and resourced.

Recommendation 5: Enhance compliance with corporate self-evaluation reporting requirements by clarifying the roles, procedures, frequency and reporting requirements for supervision, implementation support, and completion. In addition, **work with IDEV on an**

appropriate and realistic timeline for timely PCR/XSR validation in order to feed the formal annual development effectiveness report on the AfDB-funded interventions.

Recommendation 6: Enhance learning through SESP by improving (i) quality of and accessibility of SESP products, and (ii) task managers' understanding and acceptance of SESP as learning tools rather than just as accountability tools.

Table 1: SESP Reporting tools and performance

Reporting tools	Outcomes	Main issues	How well does it support the related outcome?	Proposed actions	
Public Sector Projects	IPR	Performance Accountability	Low Compliance, filing delays and candor issues. Redundancy with BTORs and no differentiation for special situations (PBOs, fragility). Managers are not sufficiently accountable for IPR quality and reliability: little validation and discussions.	Weak tool for raising issues, corrective action and performance management. IPR weaknesses permeate through PCRs, CSPs, CPPRs and RISPs. Insufficient accountability tool.	The IPR format should be updated, simplified to reduce redundancy with BTORs, and differentiated for diverse typologies of operations and country circumstances (e.g., fragility). Make Management vetting of IPRs a requirement for compliance monitoring through the dashboard.
	MTR	Performance	Compliance is low and MTRs are only occasionally used for project restructuring, which is perceived as having high transaction costs.	The tool is considered adequate but its implementation weak.	Include MTR compliance as a dashboard indicator. Requires higher degree of Mgt. accountability.
	PCR	Accountability Learning	Weak incentive structure for candor. The rating system is not conducive to effective and reliable assessment, leading to disconnects with IDEV's validation. Occasionally used as learning tool esp. for follow up projects in the same country. Excessive focus on ratings hampers learning opportunities. Trade-off between accountability and learning.	Equating project performance with staff performance leads to candor issues and reduces accountability. Learning potential is unmet as there are too few opportunities for more structured lessons-sharing events.	Reform the rating system to allow to better capture the reality and improve the guidelines for generating more solid evidence. Develop a distinct approach towards learning with the SESP as one of its inputs but relying on more adapted venues and products, a repository of information and a safe space environment.

Reporting tools	Outcomes	Main issues	How well does it support the related outcome?	Proposed actions	
NSOs	PSR	Performance Accountability	Low compliance and redundancy with ASRs & BTORs. Little discussion or validation by Mgt.	Weak tool for performance management. Low compliance reflects low accountability.	Rationalize its use, format and content relative to the other reporting tools (ASRs, BTORs). Make reporting on compliance more visible.
	ASR	Performance Accountability	Low compliance and redundant with BTORs. Lack of consistency in using common assessment criteria and rating systems between the various documents. DO ratings are not systematically validated.	Weak assessment of risks to development outcomes. Low compliance reflects low accountability.	Develop a more precise and actionable assessment of risks to development outcomes. Improve accountability through better and harmonized results reporting.
	XSR	Accountability Learning	Lenient ratings in the XSR with considerable validation backlog. Compliance issue. Projects assessed against indicators not included in the PAR or the ADOA. Few learning opportunities.	Low compliance and candor affect accountability and efficiency of reporting. Financial performance dominates over development outcomes.	Better alignment of criteria and harmonization of rating systems among the various reporting tools. Make lessons more relevant and useful.
CSPs/ CPPRs	CR	Accountability	The quality of CSPs/CPPRs is impacted by the quality of IPRs. CSPs/CPPRs completion reports are not validated by IDEV. The weight of CPPRs in assessing the quality of CSPs is questionable. CSPs/CPPRs are reviewed by the Board and are more likely to be subject to discussion and learning.	Lines of accountability between the CSPs and the CPPRs are blurred. IDEV carries out independent evaluations which facilitates accountability and learning. The current focus on the narrative, rather than on ratings, facilitates learning.	CPPR methodology and articulation with CSPs need revision. Consider validations of more CSPs/CPPRs by IDEV.
		Learning			
RISPs	CR	Accountability Learning	The quality of RISPs is impacted by the quality of IPRs. RISP Completion Reports are not validated by IDEV. CSPs are reviewed by the Board and are more likely to be subject to discussion and learning.	IDEV carries out independent evaluations of RISPs which facilitates accountability and learning even if methodologies are not harmonized with CRs.	Consider more validations of RISP completion reports by IDEV.



Management Response

Management welcomes IDEV's report on the African Development Bank's Self-Evaluation Systems and Processes (SESP). Overall, Management agrees with the recommendations which are useful in further improving the processes and tools in place to monitor and manage operations quality and portfolio performance, and sharpen learning and accountability. While many of the issues and challenges that this report highlights are covered in responses to previous evaluations and in the Bank's Integrated Quality Assurance Plan; this management response provides a reaction to this new report and sets out actions to address the specific recommendations.

Introduction

Over the years, the Bank has adopted a coherent set of policies and strategies for setting out principles, criteria and tools for increasing the quality and delivery of operations, making use of self-evaluation tools. The focus on results has also been enhanced by the introduction of an integrated project quality rating system which includes quality at entry, quality of supervision, and quality at exit for sovereign operations. Similarly, for non-sovereign operations (NSOs) tools, rating methodology and systems have been introduced to ensure quality across the project life cycle.

As already highlighted in three previous IDEV reports on similar topics, the systems require improvement, not least in relation to compliance. Since 2018, the Bank initiated comprehensive diagnostics of the quality assurance and monitoring and evaluation (M&E) tools and processes applicable to sovereign and non-sovereign operations, in order to complement work conducted by IDEV. Based on the findings of these analyses, Management prepared an Integrated Quality Assurance Plan (Quality Plan), covering five reform areas, which address many of the issues raised in this new report. As such Management notes that many of the same issues are highlighted in this report, but welcomes the focus of IDEV's recommendations on issues that are not already explicitly addressed elsewhere.

This report on the Bank's self-evaluation systems and processes presents a description of the evolution and structure of self-evaluation systems in the Bank. In practice, it focuses on self-evaluation of individual operations (as opposed to country or other strategies, or portfolio level). Within this area, the focus of the report's coverage is mainly on sovereign operations. In this respect its coverage is similar to that of the 2018 Quality of Supervision evaluation and the 2019 syntheses of the 2016-2017 PCR validation exercises, both of which have been discussed at CODE and have informed the Bank's Quality Plan. As such, it is helpful in broadly validating the directions set out in that plan.

Management appreciates IDEV's positive feedback on the strong standards and procedures underpinning the Bank's self-evaluation system, relevance of the tools and processes in place and their broad alignment with comparators' best practices. We also recognize, however, that there is considerable scope for improving the application of the established procedures, standards and norms, as well as revision of key guidance documents. At the same time, Management agrees that the observed shortcomings are primarily driven by incentives and work pressures rather than the inadequacy of templates and guidance. Many of these issues are covered in the Integrated Quality Assurance Plan, others are also being taken up in the strategic staffing review.

Management appreciates IDEV's constructive recommendations and indeed the constructive relationship that we have established on many of the issues highlighted. Management notes, however, that the report differs from that planned at inception in two ways.

First, Management would have welcomed more analysis on NSOs, Country Strategy Papers (CSPs), and the broader monitoring and evaluation system – such as roles and resources, as these are the areas less well covered in previous linked evaluations. This report includes detail in particular on specific tools used for sovereign operations. Some conclusions, where based on analysis of sovereign operations, may not necessarily be as relevant for NSOs.

Second, in terms of evidence base, the report draws on existing evaluations published in 2018 and 2019. In terms of new data, the number of projects examined on a desk review basis, (12 sovereign and 5 NSO) and the low level of responses to the survey (27 staff) are limiting. However, a significant number of additional interviews were conducted, which has enabled inclusion of staff, management and Board member perceptions.

One useful aspect in independent evaluation is the benchmarking with others. The focus of this benchmarking is on comparing what is on paper in terms of methods used, which is helpful, but information on compliance and resource issues — would also add particular value. For NSO, IDEV might consider in future consistently looking at not only IFC and AsDB but also EIB and IDB Invest.

Management stresses its overall appreciation of the evaluation, for raising or re-emphasising important issues, and the pertinent recommendations. Most importantly, the discussion with IDEV on how to tackle these challenges is constructive and collaborative. Management agrees with the recommendations made.

In order to respond most effectively to the key issues raised in the report, this response is structured as follows. The first three sections address issues around the SESP at each of: (i) project; (ii) country strategy; and (iii) corporate level. The subsequent two sections address issues that cut across the system: beginning with incentives and resources and followed by the accountability and learning functions. Finally, the response to the individual recommendations are set out in the Management Action Record.

Project Level Self-Evaluation

The bulk of the coverage of the report relates to project level self-evaluation and coverage is strongest on sovereign operations. The Bank produces Implementation Progress Reports (IPRs) and Project Completion Reports (PCRs) for sovereign operations. For non-sovereign operations the equivalents are Annual Supervision Reports (ASRs) and Extended Supervision Reports (XSRs). In this context, the report raises three important issues.

First, the report questions the rating system in place for IPRs and PCRs. With regards to the rating scale, Management agrees that the scale currently used (in which the middle two ratings do not allow for any nuance) is not ideal and is discussing with IDEV alternative approaches. It will be important to ensure that self-evaluation and independent evaluations use a similar scale and definitions. IDEV's own move from a six point to a black and white four-point scale, illustrates the same challenge. Moreover, as explained in the Management response to the 2016-2017 PCR validations, provided in 2019, Management is committed to improving the quality of PCRs and XSRs, including the rating candor. For NSOs, the recent revamping of NSO supervision tools as part of the Quality Plan's has included examination of the rating scale in use looking at each of development objectives and implementation progress.

Second, the report raises the issue of compliance, which was also raised in the 2018 Quality of Supervision report and therefore addressed in the Quality Plan. Management is seeking to address compliance issues, including through key performance indicators (KPIs). In 2020, new KPIs have been added relating to compliance with implementation reporting requirements. In addition, the Quality Plan encourages not only strong supervision (including strong IPRs for sovereign operations and ASRs for non-sovereign operations); but also, proactive implementation support and portfolio management. This is in line with IDEV's observation that the incentives to build a reputation as someone who can proactively fix problem projects need to be strengthened.

On the NSO side, it is the nature of the facility, type of investment and stage of the project that informs the portfolio officer on the level of field supervision required, this is a risk-based approach. This means that there will be times when only one field mission is carried out and others where four many be necessary. An additional dimension to consider is that in most instances, the Bank is a lender (or investor for equity) alongside other lenders. As a result, a Common Terms Agreement spells out the frequency of supervision missions. Such realities of the NSO system need to be considered when drawing broad conclusions. For environmental and social (E&S) compliance in particular, and while issues and findings related to E&S are acknowledged, it is worth emphasizing the staffing issues, which are being addressed. This is an important consideration when benchmarking with sister institutions, where the supervision teams in charge of NSOs are composed of up to 6 staff with the required set of expertise (including E&S) while the Bank has only one or two staff conducting NSO supervision.

Third, the report raises the issue of the lack of available validated ratings which management can use to aggregate results data in its external reporting. The report recognises that the reason that the Bank

has been using self-evaluation ratings for reporting in its Annual Development Effectiveness Reviews, is because the IDEV-validated ratings have not been made available in time for reporting. As discussed in the management response to the 2016-2017 PCR syntheses (delivered in mid-2019), Management very much welcomes the validations. Similarly, for XSR validations, Management looks forward to receiving validations more recent than 2014, to enable the data to be used. By providing the validated ratings within the year following PCR and XSR delivery, IDEV practices would also be brought into line with the existing Operations Manual. Management is fully committed to reporting the validated ratings for any year for which they are made available. The timely availability of validated ratings is something that both IDEV and Management want to address as soon as possible.

The report refers to staff perception of overlap between back to office reports (BTORs) and formal monitoring reports (IPRs and ASRs). The IPR/ASR can be included as annex to a BTOR, which in turn need only be one page including information not suitable for the corporate reporting, focusing especially on actions to be taken. There is also an issue of overlap with the Aide Memoire. Management is looking into clarifying and streamlining reporting requirements. For NSOs, the differentiation between BTORs and ASR has been unpacked in the new NSO Business and Process Manual.

The report in some places argues that staff have incentives to rate project performance generously, because their managers associate staff performance with project performance. In other places, it states that not enough attention is paid to project performance and compliance with project reporting standards when it comes to staff objective setting and performance review. Consistent with previous IDEV reports covering this topic, we believe that incentives are more aligned with this latter insufficient focus and compliance review situation, and the remedy should be aligned accordingly.

A general issue with self-evaluation is guarding against positive bias, especially when the same staff member designs, implements and self-evaluates – as is the case for sovereign but not NSOs. Management is looking at various options to address this, which include further segregation of duties and/or enhanced degree of review and compliance check.

Country Level Self-Evaluation

As mentioned above, the report does not include detailed analysis on country strategy level self-evaluation, which includes the Country Strategy Paper (CSP) Completion Reports Regional Integration Strategy Paper (RISP) Completion Reports and Country Portfolio Performance Reviews (CPPR). Some clarifications on the current situation may be helpful.

With regards to lines of accountability between the CSPs and the CPPRs: it is clear, the country program officer has accountability for the CPPR while the country economists have accountability for CSPs/ RISPs. Both work with the relevant country manager.

The report questions the weight of the CPPR in assessing the quality of CSPs. In fact, the only formal rating provided in a CSP completion report is that of the most recent CPPR. In 2019, a Board seminar was held to discuss the shortcomings of the current CPPR approach, and work has started to rethink it. In the context of the Quality Plan, management has committed to revise the CPPR methodology, though will do so later than planned to allow IDEV to complete an ongoing evaluation covering this topic, so the findings of that evaluation should feed into the new approach.

Management is examining the option of moving towards a dashboard approach. Such dashboards can be designed to provide a transparent overview of objective information, more informative in helping to identify the source of challenges and lessons going forward, less open to positive or other

biases, as compared with the traditional composite rating approach. This approach would not only help significantly to improve the buy-in from CSP/ RISP teams in the evaluation, but also allow for a monitoring of the evolution over time of the quality of the strategies and their implementation.

The report talks about the possibility of IDEV conducting CSP validations. To do so, it would be important first to examine the success of the pilot exercise, look at other models, and if opting for the validation model, develop an agreed methodology for CSP completion reporting, to enable the validation.

Corporate Level Use

As noted above, the main issue with regards corporate reporting relates to the availability of independently validated PCRs and XSRs.

Secondly, the report discusses the Results Measurement Framework (RMF). The RMF is due to be revised in 2020.

Besides identifying lessons for the Bank, the RMF and Annual Development Effectiveness Review reports are intended to promote transparency and accountability, enabling the Bank and its partners to check to what extent we are on track to achieve our targets. As also explained in previous management responses and the GCI VII paper, the RMF is due to be revised by the end of 2020. In this context, issues raised regarding breaking down compliance for sovereign and non-sovereign operations, as well as the realism of some of the targets are well noted. Management and IDEV have already held a first informal meeting to discuss (i) the RMF review this year; and (ii) IDEV's planned evaluation of the RMF in the same period. It will be important to ensure findings are available to inform the review of the RMF. Indeed, Management is willing to delay the RMF review to ensure that IDEV's findings (expected in Q4 2020) can be used to inform it.

An important component in terms of the Bank's ability to report corporate performance using the RMF, is the choice of indicators used at project level — currently there is lack of adherence to corporate or RMF indicators, which makes the job of aggregation challenging. As part of the revision of the Results Based Logical Framework (RBLF) committed to in the context of the Quality Plan, Management plans to strengthen guidance on the importance of using core corporate indicators where possible. At the same time, the programmed review of the RMF provides an opportunity to ensure the corporate indicators are consistently appropriate for project level reporting.

The Results Reporting System will improve monitoring and ensure consistency of results and objectives throughout the project cycle and will enhance accountability for supervision and completion. It will also facilitate more effective portfolio reviews and reporting exercises and will provide useful information to improve the design of new operations. As explained in the Quality Plan, while for SOs the system is launched and roll-out ongoing following the SAP upgrade, (and now adjustments are being made relating to remote working arrangements); Management has made significant progress in defining a similar system for NSOs. This new system is currently in development and is aimed at, among other things, harmonizing development outcomes' indicators, tracking and measurement across the NSO life cycle, and better record keeping supporting corporate reporting and indeed evaluation.

Incentives and Resources

The report rightly points out the critical roles of incentives and behaviors in the use of self-evaluation tools, including those relating to monitoring ongoing and closing operations.

Management notes IDEV's recognition that project teams and task managers are motivated to help clients deliver results and are committed to the supervision task, but the recognition that effective implementation of the SESP is affected by the

broader issue being taken up in the strategic staffing review — inadequate staff in key places, especially given the increasing number of new projects and growing ongoing portfolio. This is consistent with a similar point made regarding workload on task managers specifically in the 2018 report on quality at entry.

Management's assessment on the right-sizing and strategic staffing analysis concurs with the evaluation finding that there is a need to increase staffing and skills required to implement the Bank's monitoring and self-evaluation systems effectively. In addition to the workload of task managers, particularly in certain sectors and regions, the Bank recognizes that certain specialised support functions also need immediate buttressing. These include environmental and social safeguards, as explained in the safeguards action plan, but also expert support on results measurement and monitoring, integrated within operations to ensure proper quality check of all SESP products. The project M&E function at project level remains the responsibility of the task manager (sovereign operations) and the origination officer and portfolio officer (NSOs). The results measurement and monitoring function will support the overall system by working to reinforce quality, compliance and continuous improvement of the Bank's monitoring and self-evaluation efforts.

The Business process reengineering exercise is also focusing on simplifying the operations process, streamlining reporting requirements, and improving systems and IT. In so doing, the emphasis is on effective processes, i.e. not compromising quality as we seek efficiencies.

As explained in the Quality Plan, Management will develop a programme that rewards excellence and innovation in project design, project implementation and project restructuring. Recognising the importance, not only of good design but also, of proactive implementation - including project and portfolio turnaround successes - will support increased focus on problem solving, finding solutions and learning.

Accountability and Learning

The One Bank approach is underpinned by three cardinal principles: quality, delivery and joint accountability. The joint accountability principle is about clarifying and strengthening accountabilities for high quality delivery. The revised Presidential Directive (PD) and Delegated Authority Matrix (DAM) will clarify the roles and responsibilities and reinforce managers accountability for quality and compliance control. These roles will be further improved through operations management systems, by digitising reports development and approvals in the SAP-integrated Results Reporting System and the improved Bank-Wide Program Processing Schedule (BPPS).

Refined 2020 KPIs, which seek to find the right balance between quality and delivery have been agreed. They include new KPIs relating to delivery not only of PCRs but also IPRs and ASRs — helping to tackle the compliance with SESP requirements issue — with the KPIs filtering down by both sector and region, thus enhancing clear lines of accountability on multiple levels.

Management notes with satisfaction the report's finding that the Bank dashboards and the Results Measurement Framework are perceived to be powerful accountability mechanisms. Management plans to conduct continuous enhancement to these tools to further improve the focus on results and accountability. One important aspect is not only the

coverage of the Operations Delivery Dashboard, for example, but ensuring that SESP related indicators are discussed explicitly at Monthly Operations Status meetings.

Management agrees that the system, while conceived to serve both accountability and learning purposes, in practice tends to be more focused on accountability, at least at the corporate as opposed to task team level. This challenge of transforming self-evaluation into cross-team learning is one common area across all MDBs. Certainly, learning can be improved, but it is also important to acknowledge what is already in place, including how it can be improved.

At project level, the reporting requires efforts to capture learning: i) at quality-at-entry, appraisal specifically includes the incorporation of lessons learnt, ii) the concluding section of the IPR deals with “lessons learnt during implementation”, and iii) the PCR requires the identification of lessons for each of the four quality dimensions rated (relevance, effectiveness, efficiency, sustainability). The same goes for the revamped ASR and XSR templates. Of course, the system can be improved, for example, the importance of lesson learning is also being taken up in the ongoing work to revise the quality at entry tools or readiness review and technical peer review. At the CSP/RISP level, it has been the Bank's practice to prepare the completion report before the new strategy, as a means of identifying lessons of the past cycle to inform the forthcoming CSP/RISP.

One of the existing tools that staff can use to search for lessons related to a particular region or sector is the Evaluation Results Database (EVRD) which is run by IDEV. IPRs are also accessible to staff via the Intranet, and PINS produces analyses based on NSO reporting which is consolidated into an annual report. At a higher level, management has committed to reintroduce the Annual Portfolio Performance Review, which addresses major trends in the Bank's portfolio strengthened with broad areas for best practices and also lessons to be learnt. In addition, online courses relating to the sovereign operations tools – IPR and PCR – are close to finalization. The Operations Academy will later provide more comprehensive coverage for both sovereign and non-sovereign ecosystems.

Notwithstanding the above, there is certainly scope to introduce more learning focused events relating to different tools, but also different sectors and regions. Such events do take place but not on a systematic basis. Management is also happy to work with

IDEV's knowledge management division on such knowledge-based events. The African Development Institute is also planning a series of knowledge-based events aimed at internal knowledge sharing amongst staff.

Conclusion

Management welcomes IDEV's recommendations and the constructive relationship that is currently developing to work together to address challenges in the Bank's evaluation system. Many of the findings are also addressed in the Quality Plan, with several ongoing actions to enhance the tools, systems and processes for both SOs and NSOs. Management remains committed to improving the Bank's self-evaluation system, including its tools, processes, compliance and learning emerging from the system. The Management Action Record, below, sets out specific actions against the recommendations made.

Management Action Record	
IDEV Recommendation	Management's Response
Recommendation 1: Review the AfDB's results measurement framework and evaluation frameworks across the project cycle to ensure:	
<ul style="list-style-type: none"> ■ Internal consistency throughout the AfDB's results measurement and reporting system from operation to corporate level. ■ That there is harmonization between sovereign (SO) and non-sovereign operations (NSO). 	<p>AGREED - Management remains committed to having a strong results measurement framework and evaluation frameworks. Almost every part of the system is already under review or planned to be reviewed, from the guidance for supervision and completion to project level results framework and the corporate RMF.</p> <p>Further actions:</p> <ul style="list-style-type: none"> ■ Management has committed to reviewing the Results Measurement Framework (RMF) in 2020. The objective of this review is to enhance the framework's relevance to Bank priorities and capacity (SNDR Q2 2021). In reviewing the RMF management will: <ol style="list-style-type: none"> a. Ensure its level of ambition is realistic and informed by evidence; b. Sharpen its focus within the High 5s and with new emerging priorities – including GCI-7 and ADF-15 commitments. ■ Management will work closely with IDEV to examine potential harmonisation between sovereign and non-sovereign operations including relating to self-evaluation and validation. Since Management has already committed to re-examine both the PCR and XSR tools, in the Quality Plan, this is an ideal window to examine harmonisation (IDEV/SNOQ/PINS Q4 2020).
Recommendation 2: Review and revise, in close collaboration with IDEV, the PCR/XSR ratings methodology in use	
<ul style="list-style-type: none"> ■ Including the ratings scale and guidelines in order to improve the reliability of the ratings system for all operations, and to better align SO and NSO. 	<p>AGREED - Management has begun discussing with IDEV possible options.</p> <p>Further actions:</p> <ul style="list-style-type: none"> ■ Work closely with IDEV when reforming the PCR templates and guidelines including the rating methodology in order to provide more nuanced distinctions between satisfactory and unsatisfactory rating options (SNOQ/IDEV Q4 2020). ■ IDEV is conducting analysis designed to inform revision of the XSR guidelines to ensure they are aligned with good practice and exploring the possibility to harmonize with sovereign operations' completion reporting standards. PINS has already shared a revised XSR template with IDEV and will make amendments as per IDEV's guidance (IDEV/PINS Q4 2020).

Management Action Record	
IDEV Recommendation	Management's Response
Recommendation 3: Enhance the accountability for SESP products by:	
<ul style="list-style-type: none"> ■ Developing appropriate indicators and explicitly incorporating these indicators in the AfDB's top level corporate KPIs, in VPs and Directors performance agreements, and in staff performance discussions. The focus should be on delivery, timeliness, quality, usefulness and proactivity; ■ Encouraging candor in flagging and addressing emerging issues. 	<p>AGREED - the new 2020 top level KPIs include indicators relating to compliance for both SESP products for supervision and exit (i.e. IPR/PCR for sovereign operations and ASR/XSR for non-sovereign operations). Guideline revisions (above) will indeed reflect the importance of proactivity in implementation support.</p> <p>Further actions:</p> <ul style="list-style-type: none"> ■ Ensure the corporate KPIs are integrated into VPs' and directors' performance agreements and are embedded in the performance evaluations of staff and managers (CHHR, Q3 2020). ■ Use the Operations Delivery Dashboard to regularly track the quality and delivery of operations by complex, region and sector (SNDR Q3 2020).
Recommendation 4: In the context of the AfDB's rightsizing,	
<ul style="list-style-type: none"> ■ Review the SO and NSO staffing levels for M&E, and quality assurance activities, in order to ensure that the self-evaluation function is appropriately staffed and resourced. 	<p>AGREED - Management agrees with a need for a more systematic approach to strengthening monitoring and self-evaluation capacity. The Bank commits to increase M&E and quality assurance resources. Specifically, Management is the process of creating a new role, supporting quality and compliance for monitoring and results. This is in addition to enhanced training for task managers and portfolio officers (see recommendation 6).</p> <p>Further actions:</p> <ul style="list-style-type: none"> ■ Create five results and monitoring officer positions to support strengthen monitoring and evaluation practices, including providing expert advice to task teams, support compliance with standards and effective implementation of revised tools (CHHR/SNOQ/RDVP Q4 2020).
Recommendation 5: Enhance compliance with corporate self-evaluation reporting requirements by:	
<ul style="list-style-type: none"> ■ Clarifying the roles, procedures, frequency and reporting requirements for supervision, implementation support, and completion; ■ In addition, work with IDEV on an appropriate and realistic timeline for timely PCR/XSR validation in order to feed the formal annual development effectiveness report on the AfDB-funded interventions. 	<p>AGREED - Management agrees on the need to focus on compliance and timeliness, the revised PD and DAM will further clarify roles and responsibilities and reinforce managers' accountability on quality and compliance control.</p> <p>Further actions:</p> <ul style="list-style-type: none"> ■ Management will reinforce guidance in the Operations Manual on procedures and reporting requirements and will conduct training sessions to roll out to operations staff (SNSP, Q3 2020). ■ Management will provide strengthened guidance on reporting frequency and requirements for supervision, implementation support, and completion (SNOQ Q4, 2020). ■ Management will work with IDEV to define an appropriate and realistic timeline for timely PCR/XSR validation for IDEV to provide its validation data in time to feed into the Bank's annual results reporting. (SNDR/IDEV, Q3 2020)

Management Action Record	
IDEV Recommendation	Management's Response
Recommendation 6: Enhance learning through SESP by improving:	
<ul style="list-style-type: none"> ■ Quality of and accessibility of SESP products; ■ Task managers' understanding and acceptance of SESP as learning tools rather than just as accountability tools. 	<p>AGREED - Final SESP products are already accessible, but more in-depth knowledge work and dissemination of that work would support learning. The quality of the reports is also addressed elsewhere in this management response, in the management response to the PCR validation syntheses, and in the Quality Plan. With regards to task manager knowledge some of these items are covered in the Operations Academy Gateway.</p> <p>Further actions:</p> <ul style="list-style-type: none"> ■ Address SESP products, including their learning role, concretely in both the sovereign task manager pathway and the non-sovereign pathway (SNOQ/PINS Q1 2021). ■ Conduct learning and knowledge sharing workshops, /online learning drawing on SESP tools, including one on PCR quality, as per management response on PCR syntheses (SNOQ/IDEV Q3 2020).





Background and Context

Following adoption of the Ten-Year Strategy (TYS) 2013-2022, the African Development Bank (AfDB, or the Bank) has gone through the following major organizational restructuring, policy changes, and operational and institutional adjustments:

- Adoption in 2015 of the High 5 priorities¹¹ within the context of the TYS, leading to the development of appropriate strategies for each of the High 5s;
- Adoption of the new Development and Business Delivery Model (DBDM) in support of the High 5s; and an enhanced drive towards decentralization, including the creation of regional hubs; and
- Creation of structures such as the Delivery Accountability and Process Efficiency Committee (DAPEC) and the Technical Quality Assurance Committee (TQAC) to improve the operational processes.

The ongoing institutional changes, combined with the ADF-14 commitments to improve the quality of project design and supervision, and the performance of operations, provide an opportunity to take stock of Quality Assurance (QA) processes and examine how these processes can be optimized to promote efficiency, manage risks, and increase the likelihood of achieving sustainable development results.

Independent Development Evaluation (IDEV) has carried out an Evaluation of the Bank's QA processes, complementing the work undertaken by the Transformation Management Team (TMT) and DAPEC with the ultimate objective of increasing the extent to which the QA chain contributes to learning and development impact. Three standalone evaluations have now been completed:

- Quality at entry (QaE);
- Quality of supervision and at exit (QoS); and
- The self-evaluation systems and processes (SESP) (this evaluation).

Based on the conceptual synergies across these evaluations, IDEV delivered in October 2018 a "chapeau" report that examines the relevance, efficiency, effectiveness, and institutionalization of the Bank's QA processes across the project cycle, which also included the assessment of the compliance with the Bank's E&S safeguard requirements¹².

In addition, IDEV's recent evaluations also include:

- Independent Evaluation of the Integrated Safeguards System (2019);
- Quality retrospective report: an assessment of the Bank's quality assurance tools (2018);
- Comprehensive Evaluation of the Development Results of the AfDB Group (2016);
- Project Completion Report (PCR) validation of the 88 projects that closed in 2017 and the 49 projects that closed in 2016; and
- Independent evaluation of the Bank's Additionality and Development Outcomes Assessment (ADOA) framework (2014).

Based on existing evidence from evaluations and Management's own diagnostic study (2018), Management prepared a QA Implementation Action Plan (2019) for both the public sector and Non-sovereign Operations (NSOs), which sets the basis for the reform of the SESP, building on this evaluation.



Purpose, Objectives and Scope

Purpose

The definition and overview of the Self-evaluation Systems and Processes (SESP) are summarized in Box 1. Evaluation of the SESP builds upon the sequence of evaluations already carried out, leveraging evaluative evidence regarding supervision and quality at exit to examine how self-evaluation processes are implemented and can lead to better achievement of results.

The Bank implements both independent evaluations (the mandate rests with IDEV), and self-evaluation systems and processes (for which it rests with the Bank's Management). The two are mutually dependent. These systems and processes help the Bank to account for its investment effort, assess the quality of its portfolio, improve its performance, meet reporting requirements, learn from operational experiences, and make progress towards better demonstration of results.

The SESP are defined in different Bank documents including:

- The Operational Manual (OM), which was initially adopted in 1993, revised in 1999, and more recently in 2014. The next revision of the OM is ongoing;
- Delegation of Authority Matrix (DAM) and relevant Presidential Directives (PDs);
- Additionality and Development Outcomes Assessment (ADOA) framework for the Bank's Non-sovereign Operations (NSO); and
- The 4-level Results Measurement Framework (RMF).

The purpose of this evaluation is to support the Bank's Management and operational staff through its findings and recommendations in:

Box 1: Definition and overview

- Self-evaluation is defined as the assessment made of a project, country/regional program, and policy, sector or thematic reviews by the entity engaged in the activity.
- As described in the Theory of Change, the system is meant to support three main outcomes:
 - Performance
 - Accountability
 - Learning
- The main dimensions being assessed are:
 - Relevance
 - Effectiveness
 - Efficiency
- The main tools being used are: Progress reports; mid-term reviews; completion reports.
- Self-evaluation is accompanied by validation of certain products by IDEV such as the PCRs/XSRs or independent evaluation of CSPs and RISPs.

- Improving self-evaluation and performance of operations, and country/regional strategies;
- Improving the relevance and quality of the Bank's Operational Manual to improve the SESP function;
- Enhancing operational effectiveness and the tools for accountability and learning;
- Supporting the implementation of the new DBDM, and process engineering; and
- Informing the Board and other stakeholders on the functioning of the SESP and ways to improve it.

Objectives

The main objectives of the SESP evaluation are to:

- Assess SESP performance, focusing on their relevance, effectiveness and efficiency, in serving three main outcomes-improving performance, enhancing accountability, and promoting learning;
- Identify and assess the enablers and barriers that affect the implementation and results of the SESP; and
- Distil lessons and good practices, and formulate recommendations to enable the Bank to enhance the quality and performance (design, scope, implementation and results) of its SESP.

Key issues addressed by the evaluation include: (i) the enabling environment for self-evaluation, (ii) data collection tools and systems used to implement self-evaluation; and (iii) the use of

self-evaluation information for decision-making. The evaluation is forward-looking and offers Management a number of recommendations that can enhance the performance of the tools, methods, indicators, processes and incentives that are most likely to establish trust in the SESP and the credibility of their results.

Scope

This evaluation examines how the SESP and the ratings system are being implemented and applied to the following main outputs during the period going from approval to completion (exit).

- Public sector projects: Implementation Progress Report (IPR), Mid-term Review (MTR), Project Completion Report (PCR).
- Non-sovereign Operations (NSOs): Project Status Report (PSR), Annual Supervision Report (ASR), Extended Supervision Report (XSR).
- Country Programs and Strategies: Country Portfolio Performance Reviews (CPPRs), MTRs, Country Strategy and Program Completion Report (CSP-CR), and Regional Integrated Strategy and Program Completion Report (RISP-CR).
- Sectoral, thematic, policy/strategy papers: MTRs and CRs.

An important distinction can be made between the mandatory self-evaluation products listed above and voluntary evaluation studies, such as impact evaluations and occasional programmatic evaluations or retrospective studies, commissioned by individual business units. PCRs and XSRs are independently validated by IDEV, which also carried out, on a pilot basis, the validation of one CSP (South Africa) and

one RISP (West Africa). IDEV also regularly carries out independent evaluations of projects, CSPs (CSPE) and RISPs (Regional Integration Strategy and Program Evaluation, or RISPE) on the basis of the potential need for evaluative information (e.g., revision of Policy or Strategy when one expires). Validations and independent evaluations by IDEV are important complements for the effectiveness of the SESP but are not covered as such by this evaluation.

Key indicators of the SESP information are aggregated into apex corporate reports and scorecards for corporate accountability. These include:

- The Results Measurement Framework (RMF) and the associated Annual Development Effectiveness Review (ADER);
- Portfolio monitoring reports; and
- Reporting to the Board on progress in implementing strategies.

The evaluation covers the period 2013 - 2018, which represents a considerable part of the implementation of the TYS, the adoption of the High-5s strategies, as well as the DBDM and process reengineering

reforms. This period encompasses the issuing of the updated Operational Manual in 2014, with the findings from the evaluation informing the upcoming 2019 revision. This evaluation does not cover personnel, Board operations, control functions and Treasury operations.

Bank staff and managers, as well as members of the Bank's Board, are the primary intended audience of this evaluation. The evaluation, which is also expected to be of interest to comparator organizations, includes a comparative analysis of practices at the Asian Development Bank (AsDB), the International Fund for Agriculture Development (IFAD), and the World Bank Group (WBG).

Limitations to the evaluation relate to the availability of SESP products and documents, including monitoring reports and ratings, as filed by the Task Managers (TMs) in the Bank's system. This has been especially the case for NSO data. Data on the costs of the SESP were scanty and could only be inquired into through individual interviews, as no centralized information was available, thus constraining the cost-effectiveness analysis. Also, the low response rate to the staff survey (6 percent) has limited its representativeness and the results were only considered on an indicative basis to be triangulated with other sources.



Methodology

Theory of Change

The Bank's evaluation policy, the OECD-DAC criteria, and the Evaluation Cooperation Group's Big Book on Evaluation of Good Practice Standards guided this evaluation, which is based on a theory of change (ToC) presented schematically in Figure 1. The theory of change underpinning the self-evaluation architecture is based on the fundamental logic that a well-functioning SESP can improve:

- Performance management and how the availability of reliable information and evidence can help Management take timely decisions;
- Accountability and how the provision of key information at different levels (project, program, corporate) signals that the AfDB holds itself accountable for achieving results; and
- Learning and how the SESP can be a tool for sharing lessons and continuous adaptation.

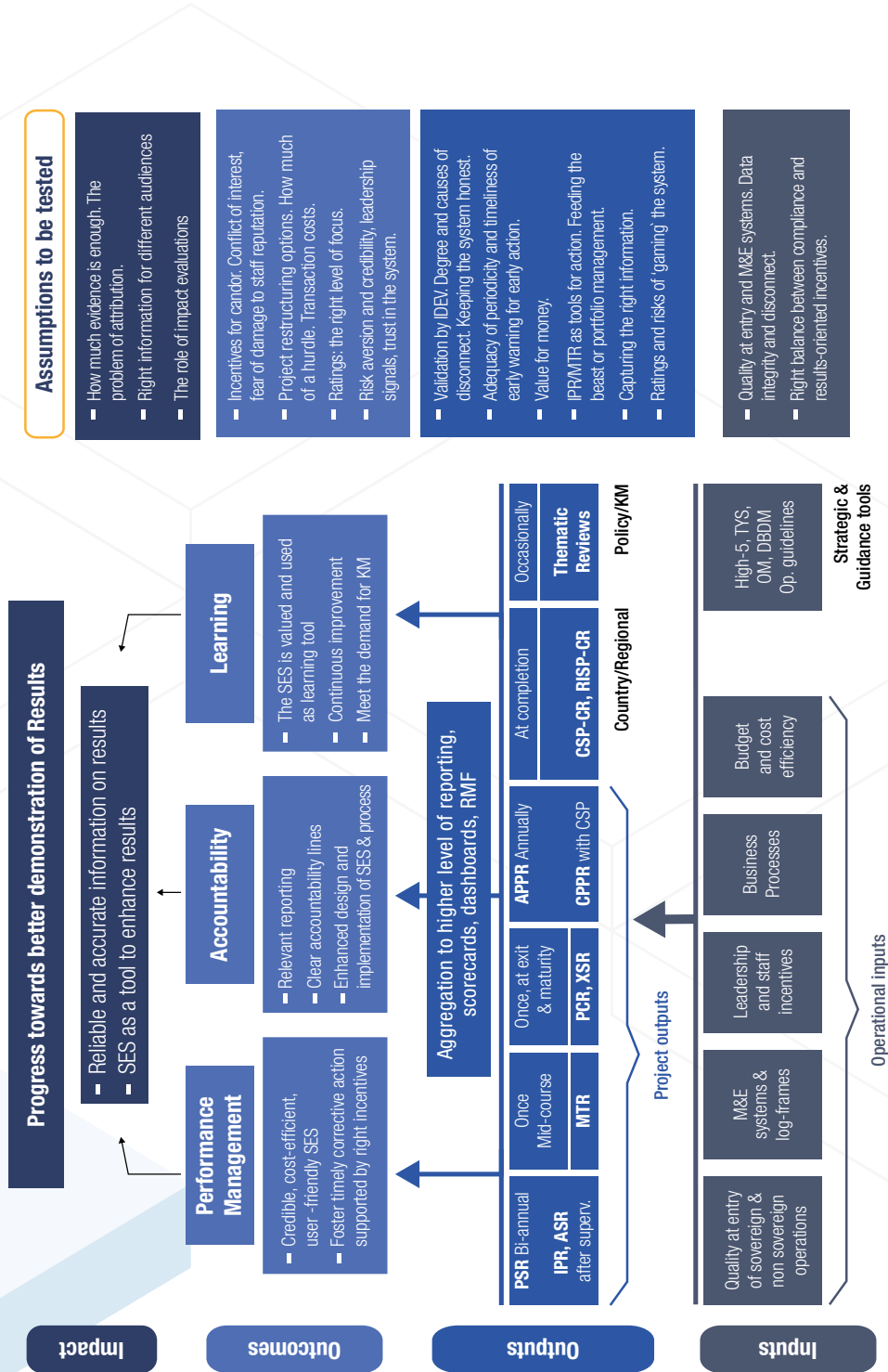
The evaluation examined the causal pathways going from the inputs into the SESP (the portfolio at entry, the M&E systems, the business processes, the leadership signals and incentive structure, and the various guidance documents), and how they influence the achievement of outputs, outcomes and impact. It also examined the links between inputs and outcomes that are ensured through the production of a number of reports (outputs) during project supervision and at closing. These reports feed into broader reporting arrangements at the corporate level. Other links between the

self-evaluation systems and other systems were also assessed, to determine how they influence the overall response culture and the incentive structure, such as: project logframes; the articulation with IDEV's own independent evaluations and 'validation' exercises (e.g., PCREN, XSREN); the commitments made at the corporate level; the Operational Manual; and other requirements.

The interface between the various systems, gaps in coverage, overlaps, relevance, periodicity, and the overall supporting environment is analyzed, also building on data from the quality of supervision and at exit (QoS) report. In examining the various causal pathways, a number of assumptions were tested to probe the robustness and credibility of the system, and to identify the weak links that could lead to recommendations for improvements. The key assumptions for the different levels of causality in the ToC cover the following:

- The effectiveness of the enabling environment and barriers to self-evaluation;
- The prevailing incentive structure and how it influences individual behaviors;
- The balance between compliance and the achievement of results;
- The adequate production, use and relevance of the project rating system; and
- Transaction costs of project restructuring and aversion to risks undermining corrective action.

Figure 1: Self-evaluation systems and processes - Theory of Change



Evaluation Questions

The overarching question addressed by this evaluation is:

“Do the self-evaluation systems and processes (SESP) support performance management, accountability and learning at the Bank”? The two underlying sub-questions are:

- How well are the SESP performing?
- To what extent are the SESP impacting on the achievement of quality development results?

Questions/sub-questions are organized in the evaluation matrix (Annex 2), as summarized in Table 2.

Evaluation Methods

The evaluation framework was structured around three main elements that represent the source of the information, as illustrated in Annex 3: (i) the Bank’s policy and guidance documents (High 5s, DBDM,

TYS, OM and Operational Directives); (ii) the main relevant evaluation reports by IDEV (CEDR, QaE, QoS, PCREN Synthesis Report); and (iii) the documents and data sources highlighted below, as well as part of the benchmarking exercise (Annex 5).

This evaluation does not duplicate previous efforts but builds on the existing base of evidence to focus more specifically on the performance of the SESP system itself. It builds on the relevant data and evidence already collected, while filling the gaps vis-à-vis new data requirements specific to the SESP. The evaluation complements the QoS evaluation. Unlike the QoS evaluation, which looks at the various components of project supervision, including from the Borrower’s perspective, the SESP evaluation focuses on the internal processes, instruments and mechanisms, with an overall objective of assessing how the application of the SESP impacts on performance, accountability and learning. While project performance relies to a large extent on the performance of Borrowers, partners and local stakeholders, the SESP is entirely under the Bank’s control and capacity to adopt recommendations.

Table 2: Summary evaluation matrix

	Performance Management	Accountability	Learning
Relevance	Are the SESP tools and processes relevant to improve performance of project/country programs?	Do the SESP generate relevant and credible information signaling that the Bank is holding itself accountable for achieving results?	Are the SESP being used as a reliable and relevant framework for learning and innovation?
Efficiency	Do the SESP provide a reliable, timely and efficient framework for portfolio management?	Do the SESP provide a reliable and efficient framework for reporting and accountability internally and externally?	Are the SESP being implemented as an efficient tool for learning?
Effectiveness & impact contribution	Is the SESP architecture being implemented as a tool to enhance performance and the achievement of results?	Is the degree of accountability exerted on the implementation of the SESP conducive to achieving results?	Have the SESP contributed to the identification and use of lesson learned?
Incentives & barriers	Are the incentives in place conducive to candid assessments and proactivity for corrective action?	Are the incentives in place conducive to exerting the right degree of accountability for the implementation of the SESP?	Is the incentive structure geared towards use of the SESP for continuous learning and innovation?

The evaluation has also built its analysis on two sets of PCR validations carried out by IDEV through the PCR Evaluation Notes (PCRENs) for: (i) 49 public sector projects closed in 2016, and (ii) 88 public sector projects closed in 2017¹³.

The evaluation methodology has followed a mixed-method approach and relied on diverse instruments to answer particular evaluation questions. Data collection methods have targeted multiple sources to enable triangulation of information. The evaluation methodology relied on the following tools (more details in Annex 3):

- **Meta-analysis** comprising literature and desk reviews of evaluations of self-evaluation systems conducted by other multilateral development banks (MDBs). This examined the common issues across MDBs, including factors affecting outcomes.
- **Benchmarking** compared various components of the self-evaluation system of comparator/sister organizations. It culls lessons of experience and good practice from AsDB, IFAD, and the WBG¹⁴.
- **Case studies** were developed around a sample of 12 public sector projects, five NSOs, three CSPs/CPPRs and three RISPs. The small sample of closed projects was selected to probe in more depth the actual application of the SESP during implementation of projects, CSPs and RISPs. Sampling was not intended to establish a base for inference on the rest of the portfolio, or to be used as a prediction tool. Instead, it was to complement already available data and provide better insight on the

effectiveness of the SESP implementation with respect to consistency and sequencing of the various reports¹⁵, and actions throughout the project cycle from the perspective of the TMs.

- **Semi-structured interviews** were carried out with 83 people, including members of the offices of Executive Directors, Bank managers and staff as practitioners and resource persons knowledgeable about the SESP across the relevant Operational Complex. Interviews were structured around the three main SESP outcomes of performance management, accountability and learning.
- **A staff survey** was circulated among staff who are directly or indirectly involved in the production or utilization of the information from the SESP. Because of the low response rate (6 percent) with only 27 respondents, the survey was not sufficiently representative to allow rigorous extrapolation and results were triangulated with other sources (QoS staff survey, staff interviews, desk reviews). Annex 7 presents a summary of the results together with the survey template.

Potential biases and conflicts of interest that could be perceived in the role of IDEV have been minimized by relying on external consultants, given IDEV's own role in the SESP. IDEV is conscious that its role might shape incentives and wants to mitigate any concern of objectivity possibly deriving from its interface with self-evaluation. The evaluation examined self-evaluation by operational staff and is not an evaluation of IDEV validation or independent evaluation.





The Underlying Factors of Self-Evaluation

The self-evaluation rationale

Few skills are more important to improving one's work than being able to step back and candidly evaluate yourself. Self-evaluation allows the raising of issues before they become too obvious to ignore and too late to correct. By human nature, people tend to be self-complacent about what they do, especially if the predominant culture in the work environment does not exercise a minimum level of accountability based on checks and balances. If the chances are that complacency will predominate, the feedback received on your own work could be warped and the self-evaluation distorted towards over-confidence.

Ignoring issues means that they will not be fixed. Self-evaluation is often confronted with deep-rooted practices that are embedded in the "culture" of the institution. While this can potentially be changed, it requires going through a revision of the predominant incentive structure. Candor does not mean having to be negative, but rather maintaining a dose of critical mindset towards oneself and objectively accept findings the way they are. Given that having to say something negative about your work can be uncomfortable, management signals need to be clear, in recognizing that project performance should not be equated with staff performance to the extent that things can go wrong for many reasons. Staff should be recognized for their willingness to raise issues as the best path towards taking corrective action and achieving results.

The Evaluation Cooperation Group (ECG) issued, in November 2018, a Practice Note: "Self-evaluation in ECG member institutions," (Annex 4) noting that if self-evaluation is to be useful it must be an integral element of and used by wider "results" systems. To this end self-evaluations should: (i) meet clear standards for quality, relevance and timeliness; (ii) be clear and well-integrated with the overall evaluation

policy to ensure management ownership and responsibilities for systems and processes; (iii) be visible and widely accessible; and (iv) rating systems must be rigorous and consistent but their application flexible.

As mentioned in the ECG Note, where self-evaluation is used as a primary evaluation tool, the main advantages are cost-effectiveness but the rigor in the validation of findings needs to be strong. An alternative to self-evaluation could be to exercise external or independent controls and mitigate the risk of conflicts of interest. This may raise the level of trust in the system but would have important cost implications and would side-line the TM who knows most about the product being evaluated. Most of the MDBs have tended to keep a combination of internal self-evaluation and arms-length/independent validation as a way of keeping the system honest (checks and balances).

Two key building blocks stand out in assessing the functioning of self-evaluation systems: (i) compliance with mandated reporting requirements to ensure that everyone follows some basic common procedures and metrics so that the various products are consistent and comparable; and (ii) candor, to ensure that the outputs are credible and can be trusted as a basis for improving the performance of the portfolio, a driver of accountability and a key motivation for learning.

Compliance

Observed compliance with project reporting requirements in the AfDB is very variable, pointing to the fact that there is little consistency in the way procedures and processes are applied, as well as the level of enforcement and accountability

exerted. The lack of compliance is an indication of how seriously staff and management take the implementation of the SESP. If compliance is weak and fails to generate the required reports and outputs, the SESP loses its foundation and credibility. This also has consequences on the quality of reporting back on KPIs to the Board and in the Management Dashboard.

Furthermore, if compliance is treated in isolation from the overall objectives of the SESP, it becomes a “feeding-the-beast” exercise, whose main objective is to ensure that the right boxes are ticked, and no flags are raised by the system. For compliance to be effective, accountability mechanisms need to work in two ways. First, they should ensure the actual production of the required outputs according to the established requirements, and second, whatever is produced should enhance the Bank's capacity to improve performance, ensure reliable reporting arrangements, and promote learning.

Staff survey respondents think that compliance can be improved through: (i) a greater dose of accountability by Managers (for instance, including comments in the IPR when clearing it); (ii) simplification and making the reporting requirement more flexible and adapted (for instance, expecting outcomes in the initial years of project implementation is unrealistic); and (iii) more coaching, mentoring and training.

Public Sector Projects

IPR compliance is particularly lacking. This is corroborated by the QoS findings (see Box 2). It should be noted that the IPR and PCR system was introduced in tandem and became mandatory as of 1st January 2013, for all public sector projects approved since 1st January 2011. For projects approved before 2011, the use of IPRs was optional and ratings could continue to be entered in SAP as before.

As reported by the Operations Committee Secretariat and Quality Assurance Department's (SNOQ) Quality Retrospective Report (2018) there is a discrepancy between actual ratings and ratings based on the appropriate application of the methodology. The IPR coverage is reported to be at 64 percent, but only 32 percent of IPRs and 37 percent of the PCRs follow the correct rating methodology, pointing to low confidence in project ratings at exit. The following shortcomings were encountered: (i) 41 percent of the IPRs reviewed have a missing rating; (ii) in many cases the justification for the rating is missing, and 15 percent of the reports provide justifications that are contradictory to the ratings themselves; (iii) 30 percent of the projects had missing baselines, which can lead to biased results reporting; (iv) just 64 percent of the IPRs were uploaded on time and 59 percent were cleared by Management on time; and (v) line managers do not verify the quality of the IPRs, as only 3 percent of the IPRs had comments and were signed off by the manager.

Compliance issues were further examined through the additional sampling of 12 public sector projects and five NSOs, including interviews with the respective TMs and investment officers. Of the 12 public sector projects reviewed, four projects were approved after 2011 had no IPRs and were not rated in SAP either. This has prevented a comprehensive analysis and understanding of the extent to which the rating of projects is a relevant tool to raise issues and to drive follow-up and corrective actions.

IPR production requirements are described in some detail in the Operational Manual of 2014. Weaknesses in the timely submission of IPRs seem to be partly related to some redundancy and overlapping content between IPRs, Aide Memoires and Back-to-Office Reports (BTORs) (this point was already noted by the QoS report). Typically, staff consider the Aide Memoire to be the key document between them and the Borrower, and the BTOR between them and Management. If IPRs are of little use and are not filed, it also means that the capacity of the ratings to be a key tool for performance management is undermined. Most of staff survey respondents

suggest some degree of merging between BTORs and IPRs, as they are both targeting an internal audience and they carry most of the overlap¹⁶.

MTRs can serve the purpose of mid-course correction in a more dedicated and targeted way than IPRs. Compliance with MTR requirements is described in the Operational Manual, and often included in the project appraisal reports and in the legal documents, which are stronger enforcement tools. MTRs are the main trigger point for entering into a discussion with the Borrower and Management on restructuring or corrective action. In principle, MTRs are mandatory, but in practice TMs decide on whether they are needed or not. Of the 52 projects in the QoS evaluation sample that were required to carry out an MTR, only 19 (37 percent) were available. Of the 12 projects reviewed as part of this evaluation, only two carried out an MTR. Meeting compliance requirements depends, among other things, on the level of attention shown by Management. If accountability is weak, staff are more likely to ignore the rules that suit them least.

In the case of PCRs, Management's own assessment indicates a compliance rate around 85 to 90 percent over the 2015 - 2017 period. Compliance is higher in this case mainly because of the need by Management and IDEV to use the PCRs in reporting development effectiveness and for validation respectively. IDEV validates PCRs, provides its own ratings in the

PCREN, and sends a synthesis report to the Board for information. Timely PCR preparation is also a KPI in the Management Dashboard. Compliance was also analyzed by the recent QoS evaluation and the major findings that are relevant for the SESP are reported in Box 2.

Non-Sovereign Operations

In the case of NSOs, assessing compliance has been problematic, as little information is available and reporting on the NSOs portfolio performance is not part of a corporate quality assurance and results system. Reporting on NSOs performance is not an explicit level-3 KPI of the RMF. The program of supervision is defined each year, depending on the type of NSO (corporate loans, project finance loans, financial institutions, direct investments, private equity investments, etc.). An NSO supervision mission can cover multiple operations. The 2018 Supervision program included 161 projects (out of a portfolio of roughly 300 operations). Of these, 65 percent had no PSRs, 20 percent had no ASRs/BTORs, and 17 percent had neither. Notwithstanding the Bank's requirement of two supervisions for each project per year, the Portfolio Management Unit (PINS) is developing a supervision framework for the debt and equity portfolio, including an early warning system to bolster the NSO supervision and monitoring activities.

Box 2: QoS Evaluation findings on compliance (public sector projects)

- 48 percent of the public sector projects reviewed had a launching mission.
- 26 percent of the projects had no supervision mission in 2017.
- 100 percent of problematic public sector projects were supervised at least once/year in 2017.
- IPRs were submitted with a delay averaging 25 days above the recommended maximum of 30 days.
- Completion rate for IPRs was not available; however, 358 IPRs were retrieved from 83 projects over six years. This compares with about 1,000 IPRs that should be available if projects were supervised twice a year.
- According to quantitative analysis, 60 percent of the IPR sample was assessed and approved by the sector manager.
- 15 percent of the projects reviewed adhered to project classification methodology and 28 percent had credible IPR ratings.
- Of the 83 projects reviewed, 52 were eligible for an MTR of which 19 were made available.
- 44 percent of PCRs due in 2017 were submitted with delays.

The 2018 annual portfolio review indicates a compliance rate of BTORs/ASRs and PSRs at, respectively, 84 and 43 percent of the active portfolio, and XSR delivery (achieved vs planned) at 58 percent. PINS reports an aggregate number for BTORs and ASRs as if they were substitutable SESP products. As reported in the NSO Semi-annual Portfolio Report (2018), the number of transactions in the NSO portfolio is increasing significantly, and thus project monitoring and portfolio management capacity should be enhanced across the NSO Department, including additional staffing and tools, as the project to staff ratio is far above the ratio of peer institutions. Furthermore, although in principle all XSRs should be validated, in reality there is a considerable backlog, as the last XSREN was produced in 2014.

Other factors

Some additional key factors need to be addressed in understanding compliance issues. With many new staff coming on board in the past few years, the Bank has lost part of its institutional memory and new TMs are not equipped with all the tools and knowledge that come with experience. The staff survey carried out as part of the QoS evaluation reported that only 44 percent of the staff received sufficient support and training for supervision. Many staff simply do not know how the SESP and the various procedures and practices work. Management's decision to initiate an Operations Academy could go a long way in addressing this issue.

Workloads are unevenly distributed across TMs, which may explain the variability of attention paid to the SESP. Some 30 percent of the TMs responding to the QoS survey declared handling more than five projects per year, with some handling up to 10. This ratio has been increasing over the years and TMs' overload represents a serious risk to quality of supervision. Furthermore, preparing new operations receives a higher priority from staff than supervision.

Candor

The difficulty in exerting candor in self-evaluation is an issue that is often admitted by staff themselves and typically attributed to the lack of a proper incentive structure. Fortunately, incentive structures can be improved so as to send the correct signals that will influence staff behavior in the right direction.

In self-evaluation, candor is applied to: (i) the narrative, which includes staff capacity to raise and formulate the issues in a way that is conducive to an objective discussion; and (ii) the ratings, which are quantitative and tend to lock the activity in a category at a particular point in time. For self-evaluation purposes, narratives are mostly used for strategy papers (RISPs, CSPs) and thematic reviews, while ratings are mostly used for projects (IPR/PSR and PCR/XSR).

Disagreements over ratings tend to be more difficult to resolve than over narratives, since by their nature ratings offer few alternative choices and less room for compromise, as the discussion is often limited to being above or below the satisfactory bar. By downgrading staff ratings in the PCRENS/XSRENS, IDEV implicitly hints to a staff candor issue, which is often not accepted by staff themselves, who argue that PCRENS are desk-based exercises that do not account for the field reality.

Public Sector Projects

A comparison between the satisfactory ratings (3 and above) assigned to the Development Objectives (DOs) and efficiency indicators in the 2016 and 2017 PCRs and PCRENS, shows significant differences leading to questioning the candor of the PCR ratings (Table 3). However, the differences in ratings decreased slightly between 2016 and 2017. Candor issues were already reported by the QoS evaluation (Box 3), with 55 percent of the staff survey respondents disagreeing that staff incentives are conducive to accurate reporting and credible scoring. The QoS desk reviews also showed

that only 26 percent of the public sector projects reviewed were fully meeting expectations on scoring credibility. As reported in the SNOQ Retrospective Report, many of the ratings of the PCRs reviewed did not provide adequate supportive evidence or justification, possibly because the PCRs are mostly based on qualitative assessments and are more

subject to interpretation. The report notes that the challenge is that PCR ratings are neither impartial nor independent, as long as the TM or a consultant working under his/her supervision drafts the completion report. This leads to inflated ratings in the self-evaluation report without supporting evidence.

Box 3: Candor: Highlights from the QoS evaluation

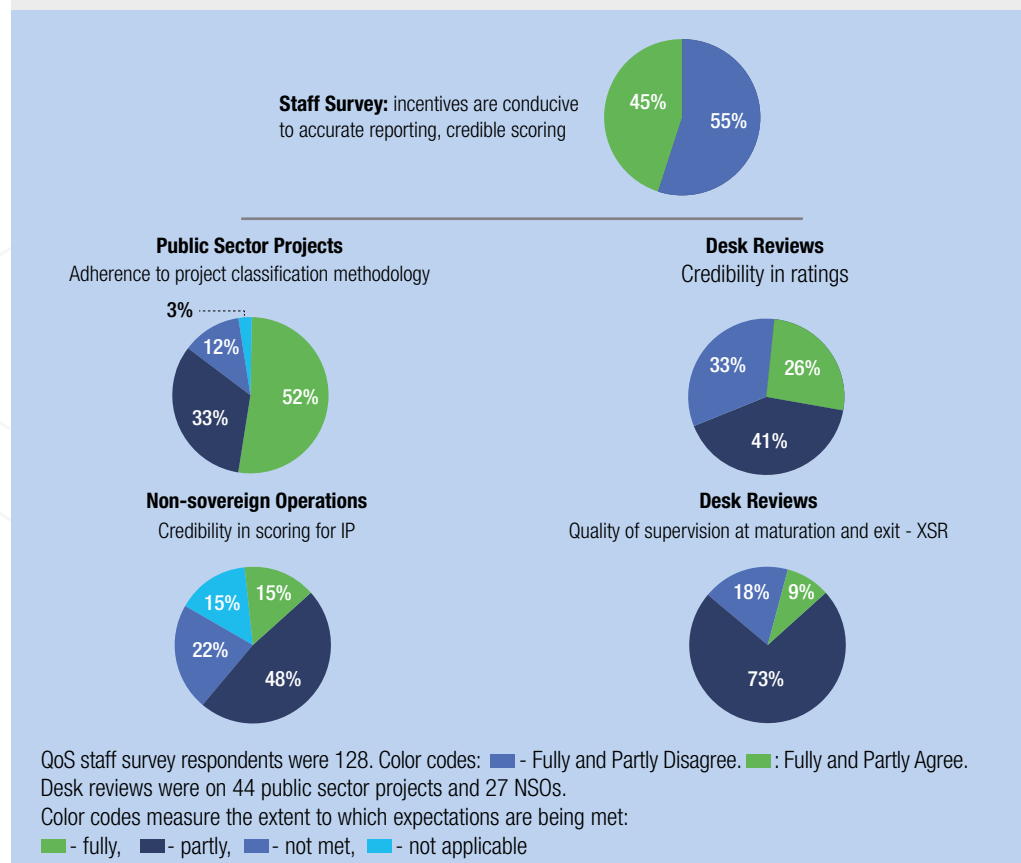


Table 3: Projects with satisfactory DOs and efficiency indicators (3 and above)

	2016 cohort (49 projects)	2017 cohort (88 projects)
DOs		
■ PCR	88%	91%
■ PCREN	53%	63%
Efficiency		
■ PCR	65%	57%
■ PCREN	41%	43%

On the other hand, interviews carried out indicate that TMs attribute most of the difference in ratings to the lack of sufficient context and in-depth information by the IDEV validation. Furthermore, they see project ratings as a reflection of their own performance. This is a sign that the Operational Manual recommendation that ratings be applied through a team-based approach has not really been adopted and that ratings remain to a large extent the exclusive responsibility of the TM who takes the entire onus on himself/herself and tends to personalize the assessment being made.

Non-Sovereign Operations

In the case of NSOs, candor seems to be more challenging than for public sector projects. Lack of supporting evidence and available data is a recurrent issue. Information available in the PSR is scanty, duplicates the BTORs, and is of uneven quality. The five NSOs reviewed independently as part of this evaluation all raised candor issues (Box 4). The QoS sample of 27 NSOs (Box 3) shows that only 15 percent of the projects reviewed met candor expectations, while 70 percent met them only partly or not at all.

PSRs do not include much information on E&S issues or the status of implementation of mitigation measures, which makes it difficult to aggregate the information in the ASR. This resonates with the analysis undertaken for 56 financial intermediary operations conducted for

the ISS Evaluation¹⁸. DO ratings seem to be the least credible as reported in the QoS report. In particular, Lines of Credit (LoC) lack consistent metrics for measuring and reporting development impacts, as also reported in the LoC evaluation¹⁹. With respect to the XSRs, PINS is collaborating with IDEV in revising the XSR guidelines to ensure they are aligned with good practice and also exploring the possibility of harmonizing them with the results framework of the public sector projects. Finally, as noted in the NSO QA Action Plan, a critical area of compliance is the clear definition of a process that defines the time of early operating maturity and timing of the XSR, depending on the NSO instrument, as the timing can also play an important role in the accuracy of the DO ratings.

Managing Ratings

The advantage of ratings is that they are easy to report on, compare, average and aggregate. They can help to trigger action and assess the attainment of targets through relatively clear-cut attribution. The disadvantage of ratings is that they can be seen as threatening, rigid, evaluative of staff performance, or not sufficiently nuanced to reflect real life situations. The ECG practice note of 2018 (Annex 4) observes that self-evaluation systems may elect not to apply ratings in some instances and should be sufficiently flexible to accommodate management and evaluator judgment.

Box 4: Key issues observed in reviewing the sample of five NSOs¹⁷

- Outcome indicators in the Project Appraisal Report's (PAR) logical framework and the Additionality and Development Outcome Assessment (ADOA) ratings are at times imprecise and lack baselines, and there is frequently lack of symmetry between the objectives/indicators in these two documents.
- Exceedingly lenient ratings in the XSR, at times with projects being assessed against indicators that were not included in the PAR or the ADOA.
- Tendency to rate projects' outcome Satisfactory despite evident shortcomings in the achievement of outcomes.
- Imprecise definition of business success.
- Considerable deficiencies in the financial and economic analysis.
- Inadequate attention to E&S safeguards during supervision.

Table 4: Satisfactory cut-off rating and disconnect

	2016 (#49)			2017 (#88)		
	PCR	PCREN	disconnect	PCR	PCREN	disconnect
2.5 and above	98%	78%	20%	97%	85%	12%
3 and above	76%	31%	45%	82%	47%	35%

Public Sector Projects

An important element of the Bank's ratings methodology is the use of simple averages of sub-ratings and simple averages of dimensions (relevance, efficiency, effectiveness and sustainability) to determine the overall rating of the project in the PCR. Because this process generates decimal points, ranges are then used to attribute the final rating as an integer number within the current 4-point scale²⁰. Accordingly, the current definition of Unsatisfactory (U) is a rating between 1.50 and 2.49, and the definition of Satisfactory (S) is a rating between 2.5 and 3.49 (see Box 5). It is therefore relevant to further examine whether these ranges can affect the overall assessment by qualifying as satisfactory projects that are not meeting all satisfactory criteria (3-rating), or more precisely those rated between 2.5 and 2.9. The analysis shows that the share of projects falling into this category is 18 percent of the PCRs and 41 percent of the PCRENs. Hence, a large portion of projects is rated Satisfactory overall even though not all dimensions are rated satisfactory.

PCRs are rated both by staff and subsequently by IDEV through the validation process. Sometimes, this generates divergence of views and a "ratings disconnect"²¹. The current rating methodology based on rating ranges can generate a different picture, depending on how Satisfactory is defined. As shown in Table 4, if it is defined as 2.5 and above, 97 percent of the projects are considered to be Satisfactory or better in the PCRs, which is questionably high by all comparable standards. In this case, the disconnect with IDEV is 20 percent in the 2016 batch and 12 percent in the 2017

batch. However, if Satisfactory were to be defined as 3 and above, satisfactory projects decline to 76 and 82 percent in the PCRs for 2016 and 2017, respectively, while the disconnect is much larger, at 45 and 35 percent, respectively. Such big differences are due to the large number of projects located in the 2.5 to 2.9 range and raise questions regarding the credibility of the system²².

Three of the comparators (WB, IFC and IFAD) do not use the averaging methodology or the ratings range system to assign the final ratings, but rather a judgment-based approach. For example, the WB uses a "tree view" with a cascading system of 4 ratings (High, Substantial, Modest, Negligible) applied to each dimension, to reach a 6-point scale overall project rating (from Highly Satisfactory to Highly Unsatisfactory) (see Annex 5, Table 17, under WBG). The Inter-American Development Bank (IDB) puts a higher weight on effectiveness, and does not allow projects lacking relevance and effectiveness to be rated positively. One should be careful in comparing the AfDB's performance with the one comparator institution that does not use the range methodology, as the cut-off rating for satisfactory

The way ratings are structured can have an important effect on the incentive for candor. Contrary to other comparator agencies (IFAD, the WBG, IDB) that have adopted a 6-point rating scale, the Bank uses a 4-point system. Admittedly, a 6-point rating system allows to capture situations that are more middle-ground between S and U, such as "partly" or "moderately" S or U. This would cater for a potentially large number of projects for which unanticipated issues have inevitably emerged in the course of implementation and that could not be fully

addressed during supervision. The review of the 137 PCRs/PCRENs reveals that the number of projects falling into an intermediate group between 2.1 and 2.9 was 20 percent for the PCRs and 57 percent for the PCRENs (Table 5). In these situations, it is often difficult to shoehorn project performance into the S or U category. Given the stark choice to be made between S or U, staff would typically lean towards the more positive rating, while IDEV would tend towards the lower rating, creating the conditions for a bigger divergence. Ratings differences between the top half and the bottom half of the scale can also appear smaller depending on the choice of words, even on a 4-point scale²³.

The DO rating is extracted from the last IPR to feed into the PCR to assess progress made as follows:

- Highly Satisfactory (4): Both outcomes and outputs are highly satisfactory.
- Satisfactory (3): Both outcomes and outputs are at least satisfactory.
- Unsatisfactory (2): Either outcomes or outputs are unsatisfactory.
- Highly Unsatisfactory (1): Either outcomes or outputs are highly unsatisfactory.

Table 5: Number of projects within rating ranges (out of 137 projects)

	Rating range 2.1 - 2.9	Rating range 2.5 - 2.9
PCR	28 (20%)	25 (18%)
PCREN	78 (57%)	56 (41%)

Box 5: The unintended consequences of the ratings system

The PCR rates 11 criteria under four dimensions: (i) the relevance of the operation from design/approval to completion; (ii) the effectiveness in the actual achievement of results (DO); (iii) the efficiency in the use of resources for achieving results; and (iv) the sustainability in the continuation of results after the project period. The performance of the Bank, the Borrower and others stakeholders are also assessed but do not form part of the overall PCR rating.

The score of each dimension is the average of the sub-criteria, and the overall score of the project is the average of the scores of the four dimensions. All criteria are equally weighted. The rating scale only allows for integer numbers from 1 to 4, but the averaging exercise inevitably leads to scores with decimal numbers. Therefore, to be able to bring ratings back in line with the original scale, the PCR preparation manual defines ranges: 1.00-1.49 (1-Highly Unsatisfactory); 1.50-2.49 (2-Unsatisfactory); 2.50-3.49 (3-Satisfactory); 3.50-4.00 (4-Highly Satisfactory).

The Satisfactory range (2.50-3.49) can be problematic to the extent that projects in the 2.50-2.99 range will eventually be categorized as 3 even if there are by definition some areas that are less than satisfactory. As an example, if all dimensions of a project resulted in an average score of 2.5, the project would still be classified as satisfactory.

The four dimensions do not carry the same importance in assessing the achievement of results and hence weighting them equally can affect the results. Most projects present high scores on relevance arguing that the project is in line with Borrower and Bank strategies (the average relevance rating of the 88 PCRs of projects that closed in 2017 is a high 3.7). This skews the overall average in a positive direction. The DO could be rated Highly Unsatisfactory and the project would still be Satisfactory thanks to the average system. The DO, which arguably is the most important criterion in terms of delivering outputs and achieving outcomes in a combined way, accounts for only one quarter of the overall score.

Other MDBs give a higher weight to effectiveness (e.g., 40 percent in the case of IDB, vs. 20 percent for the other dimensions and 50 percent for PBOs, where efficiency is not being rated).

		Outcomes rating			
		4	3	2	1
Outputs rating	4	HS	S	U	HU
	3	S	S	U	HU
	2	U	U	U	HU
	1	HU	HU	HU	HU

In reality, the majority of the PCRs reviewed present a mixed picture in terms of outputs and outcomes but typically the TM leans towards the satisfactory rating even if the project did not meet the expectations and the specifications of the satisfactory definition.

An analysis of the pros and cons of a 4-point vs. a 6-point scale has been undertaken by IDEV in the context of CSPs²⁴. The main advantage of a 6-point system is to mitigate for the stark separation existing between S and U, which is often perceived to be too wide, as conveyed by staff interviews. However, even a 6-point system will tend to classify performance of most interventions under “Moderately” or “Mostly” Satisfactory, which ultimately will be aggregated together with the Satisfactory and above categories as being “above the bar”. This is evidenced by the case of IFAD and the World Bank. Comments received from the Bank’s Evaluation Reference Group (ERG) suggested to have a simple 3-point scale of “yes”, “no” or “to some extent” arguing that the more the points the more complex it becomes and the more difficult it is to assess performance adequately. However, a 3-point scale would be against the Good Practice Standards established by the ECG of MDBs, of which the AfDB is a member.

In 2017, IDEV moved away from the 6-point scale and adopted a 4-point scale for the evaluation of CSPs (consistent with the public sector projects rating scale) on a pilot basis for two years. In doing so, IDEV noted the risk that evaluators would give the benefit of the doubt and most ratings would fall on the satisfactory side of the scale, creating tensions between IDEV and Management, and between

Management and the Board, as the distance between S and U may be too wide and cause a loss of granularity, and be detrimental to learning. Two years later, the risks proved to be valid. The same risk can also be applied to the PCR situation, with the difference that in this case PCR validations by IDEV resulted in increased levels of disconnect. The 2016 Synthesis Report indicated that the current 1-4 scoring system does not allow for very nuanced evaluations.

Any rating system will have pitfalls and, regardless of whether a 3, 4, 5 or 6-point rating is adopted, it may not lead to significant improvements unless other measures are put in place for improving the reliability of the system and the implementation of more rigorous procedures for the generation of the required evidence. This notwithstanding, the rating scale would still be worth a deeper discussion as, in the opinion of almost all the staff interviewed, the rating scale and methodology need revision.

The disconnect is not reported in the RMF, even though a target is now included on the “number of operations independently rated as satisfactory and above at completion”. In general, a high rate of disconnect suggests problems in the M&E system and in the SESP, often reflecting unwarranted optimism regarding project performance. Such risk, however, does not seem to represent a major incentive for staff to rate projects with more candor, possibly because the validation ratings are not receiving much attention.

According to the PCR Guidance Note, the PCREN ratings are those to be used officially and they supersede the PCR ratings in case of differences. This is consistent with the practice of comparator institutions. In reality, this is not the case, as PCR ratings (in the absence of PCREN ratings) are used as proxies for the official ratings instead. This has to do with the difficulty of having PCRENs ready in time²⁵, particularly when dealing with separate reporting

processes such as the RMF and the ADER report, which are fed by the project ratings. This also means that, typically, portfolio performance may present a relatively more positive picture than the reality and may be over-rated by the size of the disconnect. For instance, the RMF indicator on “operations that achieved planned development outcomes” uses the PCR rating. Starting in 2018, IDEV reduced the number of PCRs being validated to a sample of 65 in order to cope with resource and time constraints for feeding into corporate reporting.

Indicators pertaining to Bank and Borrower performance have a substantial disconnect, when considering the 2016 and 2017 cohorts of PCRs validated by IDEV (Table 6). Even if they are not included in the calculation of the overall project rating, they generally suffer of candor issues. This stems from the fact that asking staff to self-evaluate Bank performance is easily equated with staff performance, and the performance of their clients, with whom the Bank is striving to maintain a good relationship. Even in situations where the project was clearly unsatisfactory or worse, staff rated Bank performance as satisfactory. Of the 137 PCRs reviewed, the Bank's performance was rated unsatisfactory (below 2.5) only twice. This contrasts with IDEV's validation, which rated Bank performance as unsatisfactory in 30 cases. The disconnect is also present in the case of Borrower performance (22 of the PCRs vs. 45 of the PCRENs).

Non-Sovereign Operations

In the case of NSOs, the PSR, ASR and XSR evaluation dimensions are shown in Table 7. The SESP are not fully consistent in using common evaluation criteria and rating systems between the various documents, causing a lack of clarity

in the assessment process. The range of the various ratings for the different dimensions and subdimensions varies from 1-4, to 1-5 and 1-6, all with different definitions. The only harmonized ratings are those of the XSRs and the XSRENs.

The Bank's XSR preparation guidelines of 2011 are outdated, especially when it comes to defining a process for selection of NSOs. The link between the ASR and XSR documents, especially the consistency of the methodology of assessing and reporting on financial performance and impact on the Bank's Sector and Corporate Portfolio, is weak. A comparison between ex-ante (per Board and ADOA) and ex-post (XSR) is made by the XSR team using DO and additionality ratings in numbers and narratively for KPIs. The lack of symmetry in the use of criteria and ratings makes the assessment of development outcome across the project cycle difficult. As a consequence, the NSO self-evaluation is not based on a consistent methodology for assessing performance, accountability and learning. Management is committed to addressing these issues under the recent NSOs Quality Assurance Action Plan (September 2019), notably by revising the guidelines and templates for ASRs and XSRs.²⁶

Enforcement and Incentives to ensure Compliance

Understanding what drives staff behavior is essential for effective SESP. A combination of enforcement to ensure compliance and incentives will work in tandem to determine staff response. What works best for enforcement purposes is to make compliance with the different requirements public and transparent through a central IT-supported

Table 6: Bank and Borrower performance (out of 137 projects)

	PCR	PCREN
Unsatisfactory Bank performance	1%	22%
Unsatisfactory Borrower performance	16%	33%

Table 7: NSOs evaluation dimensions

PSR				
Implementation Progress	Commercial Viability		Development Outcome & Additionality	
ASR				
Operational/financial performance	Implementation Progress	Development Outcome	E&S sustainability	Early warning
XSR				
Development Outcome <ul style="list-style-type: none"> ■ Project business success ■ Economic sustainability ■ E&S effects ■ Private sector dev. 	Investment outcome <ul style="list-style-type: none"> ■ Equity ■ Loan 	AfDB Work Quality <ul style="list-style-type: none"> ■ Screening, appraisal & structuring ■ Supervision & administration 	AfDB Additionality	

repository of the information that can be accessed in real time. This is what has been missing so far in the Bank. It was difficult for the evaluation team to access the information and it was not always clear whether the different outputs were simply not produced, or whether they were not easily accessible through the system. Management commitment to roll out the Results Reporting System (RRS) and the ongoing use of the delivery Dashboard go in the right direction. By raising flags automatically in the system and making them visible to Senior Management in real time, the Dashboard works as an alert system and an effective pressure point, even if it is currently focused on disbursement, procurement and implementation delays, and not on development effectiveness.

Independent or arms-length validation has a significant influence over behaviors, and is an effective tool to improve candor and keep the system credible. In addition to what is carried out by IDEV on Completion Reports, internal mechanisms are also possible. Private sector operations through PINS, which is detached from the originating unit, implements a monitoring function with responsibility for reviewing and consolidating portfolio data, including XSR preparation for the projects under its responsibility and supporting those carried out at the decentralized level based on IDEV guidelines and ADOA reporting requirements. The ongoing revision of the XSR template and guidelines will enhance this process. Public sector projects have also recently

instituted a new Implementation Management Team in each Region in Regional Development, Integration and Business Delivery (RDVP), charged with providing support for effective project management and completion through results-oriented oversight, portfolio performance in collaboration with Country Managers, disbursement, legal, procurement and other service functions. It drives the actualization of the Project Implementation Plans for the achievement of project deliverables and KPIs.

Internal audits also play an important function in the mix of enforcement and incentive mechanisms, and in helping to keep both the internal and external validation functions in check vis-à-vis compliance requirement. The delivery of KPIs is a key control measure for compliance and performance.

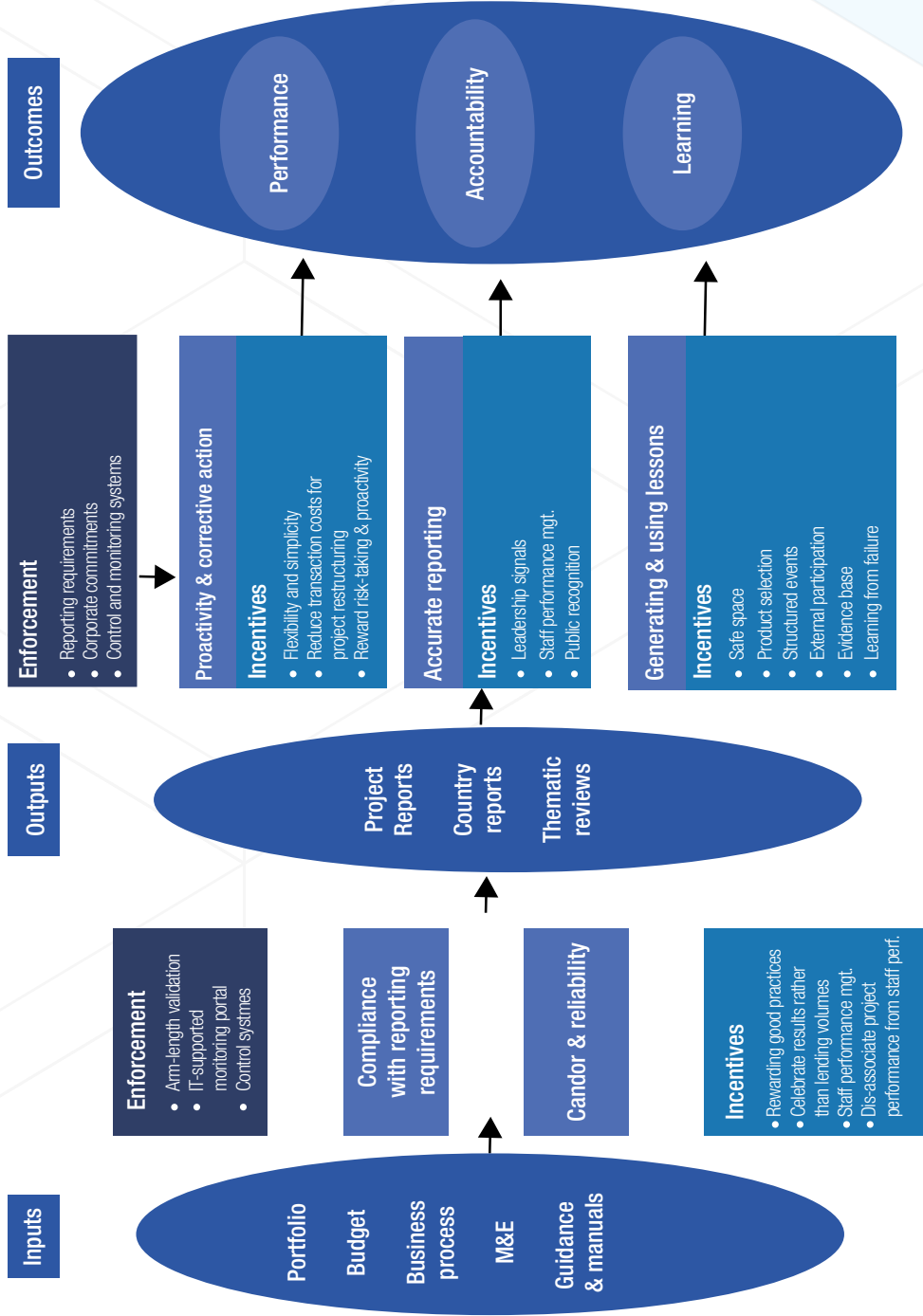
The incentive structure presents some key weaknesses that would need to be addressed to influence behaviors in the right direction: (i) if project performance is equated with the performance of staff, the fear of damaging one's reputation in case of poor results will inhibit candor; (ii) there is insufficient public recognition and rewarding of good practices related to SESP implementation, M&E systems, proactivity in taking corrective action, etc., while such recognition could send a strong signal that Management cares about staff delivering results (and not just loan approvals);

(iii) budgets for SESP products seem to be on the low side compared with sister organizations²⁷; and (iv) as reported during staff interviews, there is no systematic discussion on the achievement of results during the staff performance evaluation.

The right mix between enforcement and incentives will need to be worked out to ensure that achievement of the SESP outcomes is

not hampered by the same instruments being used. An excessive focus on ratings may lead staff to look at self-evaluation mostly from an accountability point of view, undermining the capacity of the SESP to be a tool for learning. If ratings are biased towards over-optimistic views, it will hamper the SESP capacity to be a tool for corrective action.

Figure 2: SESP incentive structure



Main SESP Characteristics and Outcomes

This chapter answers the question of whether the SESP architecture and its instruments are relevant, effective and efficient in the way they are being implemented, to enhance portfolio performance, accountability and learning. It addresses this question by looking at the key elements needed for a well-functioning SESP.

Performance Management

M&E systems

Public Sector Projects

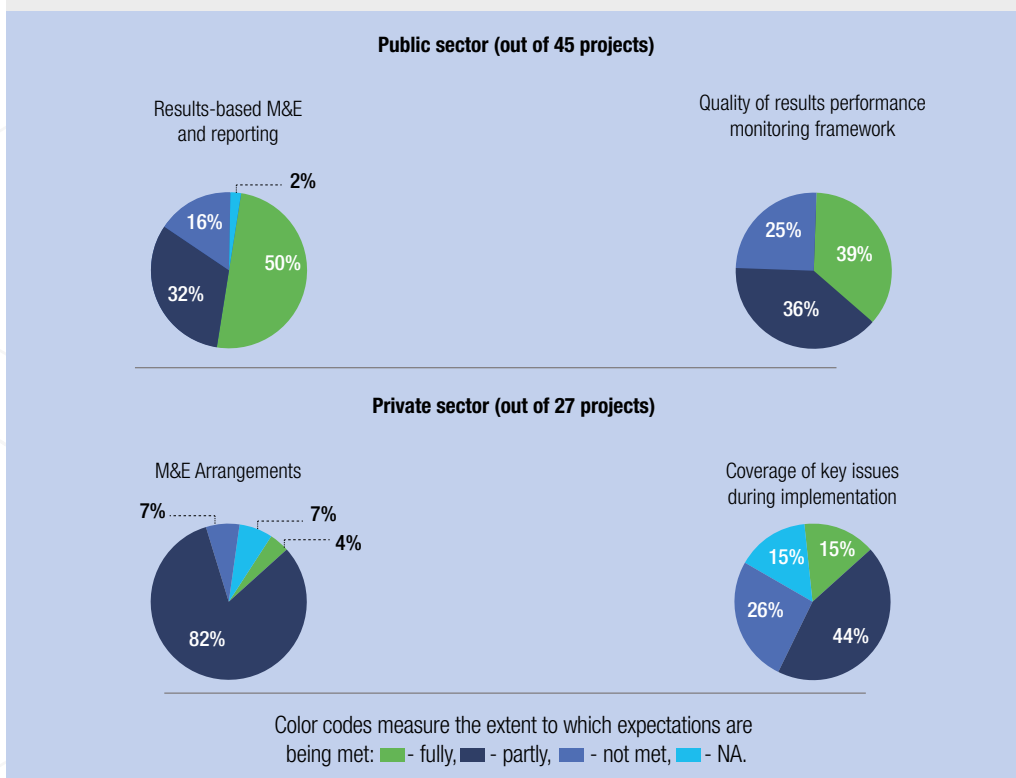
M&E systems are the building block of any self-evaluation system, and they are also the weakest point as reported in both the 2016 and 2017 IDEV Synthesis Reports. The analysis of the 49 PCRENs of 2016 noted that 20 projects did not have enough information to be able to score the M&E at implementation and of the rest, 22 of the 29 projects that could be rated were considered to be unsatisfactory or worse. The average score was only 2.07 for design and 2.02 for implementation. The gender indicator scored only 1.97 and, of the 33 projects for which a rating was available, 24 were considered to be unsatisfactory or worse. The 88 PCREN of 2017 noted a 41 percent satisfactory assessment of M&E overall with a rating of 2.7 at design and 2.4 at implementation. The gender criterion was not rated or was unsatisfactory in 49 cases. It appears that there is clearly a need to improve the M&E design, especially in the gender dimension.

The M&E framework with indicators (2016 - 2020 update) provides operational guidance for reporting on

climate change issues at project or CSP/RISP levels. However, the status of E&S safeguards is missing in portfolio flagship reports. The level of attention devoted to safeguard issues in the SESP during implementation compares poorly with the attention devoted to these aspects during project design. This finding is corroborated by the ISS evaluation, which highlights understaffing as the most significant constraining factor to the implementation of the ISS, and that the current staff and long-term consultants are not able to ensure that 50 percent of high-risk projects are properly supervised.

Weaknesses of M&E systems are often reported at the design stage already. These are carried over into implementation, but they are typically more difficult to fix during supervision, as evidenced by the fact that ratings do not improve from design to implementation. The lack of explicit links to outputs and outcomes, weak logical frameworks, delays in submitting progress reports by the borrower, as well as difficulties in recruiting and retaining M&E experts, were noted. There are few M&E specialists in the Bank, and none in the regions, which hinders the effective delivery of SESP outputs.

Findings from the 2016 and 2017 cohorts of PCRENs highlighted weaknesses in terms of information gaps with respect to evidence justifying ratings, which resonates with indications of absence of quality control mechanisms. Many ratings were downgraded by IDEV's validation, not necessarily for failing to meet the target but for lack of evidence that would support a particular assessment. This feedback resonates with the QoS desk review (Box 6) also showing mixed results with respect to overall performance of the M&E system and results monitoring frameworks.

Box 6: Mixed results on the performance of the M&E system and results monitoring frameworks

M&E issues are not new to the AfDB and have also been identified in comparator institutions. The TMs interviewed stated that significant efforts were being made to implement and use a sound M&E system but that more measures were required to encourage adequate use by borrowers. The limited capacity of borrowers to make adequate use of M&E systems constrains the use of the full information by the TMs when preparing IPRs. Data quality at the source remains an issue and affects the credibility of the whole system. Measuring and monitoring performance has often been an elusive target, especially when outcome indicators are difficult to quantify, making the assessment more arbitrary. If results reporting is to receive the required priority, the Bank and the Borrower must pay proper attention to the integration of: (i) baselines, (ii) logical frameworks, and (iii) M&E systems²⁸.

Non-Sovereign Operations

The private sector department follows a risk-based approach to the supervision of NSOs, which is performed by portfolio officers and a credit risk team that monitors the project credit risk performance. Overall, its approach is considered to be relevant and aligned with comparators. However, the staff interviews indicate that the originating team may not learn much from its experience in project supervision because of the lack retrospective feedback on the results of their transactions. In general, the stronger focus on the bankability of the project may take incentives away from providing the right level of attention to the development impact of projects.

M&E arrangements and coverage of key issues during supervision have been difficult to analyze for lack of information and limited availability of PSRs. Providing evidence that would support the project development rationale and causality between project activities and impact was found to be problematic in NSOs (as also reported in the QoS). Clients' reports on development outcomes are not assessed or validated by the NSO portfolio managers, making it difficult to ascertain their reliability. To some extent, the nature of NSO clients makes them more concerned about financial returns than development objectives.

It was also difficult to assess the covenant compliance in the NSO consolidated portfolio report with respect to Annual E&S Monitoring Reports, which are required to be prepared by clients and cover, among other things, E&S criteria, health and safety monitoring indicators, social programs, greenhouse gas emissions and energy efficiency. Monitoring Report findings are not systematically followed up and not sufficiently reported on in PSRs, ASRs and XSRs. As also noted in the ISS evaluation, E&S supervision of private sector projects is lagging despite recent ad-hoc efforts to improve the templates for some NSOs. However, the level of coverage of E&S information in XSRs was better than in the supervision reports analyzed.

The NSO consolidated portfolio management and monitoring report, prepared jointly by the various departments and PINS, is mainly focused on financial risk assessment, value addition to the Bank's income, and return on investments, but does not include an overall assessment of the achievement of development outcomes. The recent diagnostic study of the existing results system, undertaken by IDEV, revealed important gaps in terms of the monitoring of achievements in private sector interventions, including potential inconsistency between DOs set at origination and those that are tracked during implementation. Half of the financial sector projects examined by the QoS (11 projects) had unsatisfactory supervision reports and the other half were partly satisfactory. To address this challenge, the Bank has put in place a new initiative, led by the SNOQ and PINS departments, to harmonize the

results reporting requirements and ensure pro-active management of development results²⁹.

In the case of Lines of Credit (LoC), the problem stems from the poor definition of how they will be used and their likely impact on DOs. As reported in the LoC evaluation, the key success measurement is the timely reimbursement of the sub-loan and profitability for the lender. But the effectiveness of LoCs at the end-beneficiary level is often questionable because evaluation criteria are missing and the attribution of development results to LoCs is difficult.

Raising issues in a timely fashion

Public Sector Projects

TMs are required to provide information on project progress following a field supervision mission or desk review within 30 days, through an IPR, which is subsequently approved by the line manager. The timeliness in reporting after supervision missions is critical for the prompt raising of issues and the taking of remedial action. One of the main functions of project supervision is the follow-up and resolution of issues identified in previous supervision missions through the SESP.

Deficiencies have been identified on all the points above: timeliness in submitting IPRs after a mission is poor, with an average time of 55 days, and only 53 percent of supervision reports are submitted on time against a target of 70 percent³⁰. A review of the supervision missions carried out for 83 public sector projects from 2013 to 2017 by the QoS evaluation showed that the percentage of submitted and unassessed IPRs increased from 20 to 45 percent. It was also reported that 39.8 percent of projects identified as problematic remained flagged for a long time (2015 - 2016) and this rose to 49.7 percent in the following year (2016 - 2017). Interviews carried out as part of the SESP evaluation indicated little reliance on the IPR documents, as they are typically approved automatically by the Manager with no discussion with the TM and little validation. This does not necessarily mean that issues are not raised through other means,

such as the BTORs, but it does show that the IPR, as a SESP tool, is not serving its function.

The limited use of IPRs and their ratings for the purpose of addressing issues and taking corrective action may be a missed opportunity to the extent that Implementation Progress (IP) is a good predictor of overall project performance. The correlation rate between IP and the overall project rating for the 2016 and 2017 PCRENS' cohorts is 0.67 and 0.61, respectively. Addressing implementation issues as they emerge and in a timely fashion is one of the main expected contributions of the self-evaluation system to portfolio performance and the achievement of results. If IPR ratings are not reliable, not comparable, and not timely, the willingness to act on them is unlikely to be very high. Of the 12 projects reviewed as part of this evaluation, five had not been rated, whether in SAP or in the IPRs (including four projects approved after 2011, when the IPR system was introduced).

Interview responses for the SESP evaluation indicated quality control gaps in the progress reporting process. It was reported that, while supervision teams raise key implementation issues in Aide Memoires and BTORs, in most cases these issues are either not adequately addressed or not addressed in a timely manner. The tendency not to give projects a poor score seriously limits the opportunity for early identification of issues, which is when they need to be addressed, before they become too entrenched and difficult to resolve. The IPR is not considered a decision-making tool but rather a compliance tool (administrative requirement) with insufficient accountability since they are not much discussed nor systematically validated by Management.

The quality of IPR ratings is even more important because the IP and DO ratings of the last IPRs are exported into the PCR. IPR ratings are also used to assess portfolio performance through the CPPR, which is a key component of the CSPs. Hence the quality of the IPRs also permeates into CSPs. Moreover, IPRs are not adapted to program-based operations (PBOs), which follow a different business model and for which a different format would be required.

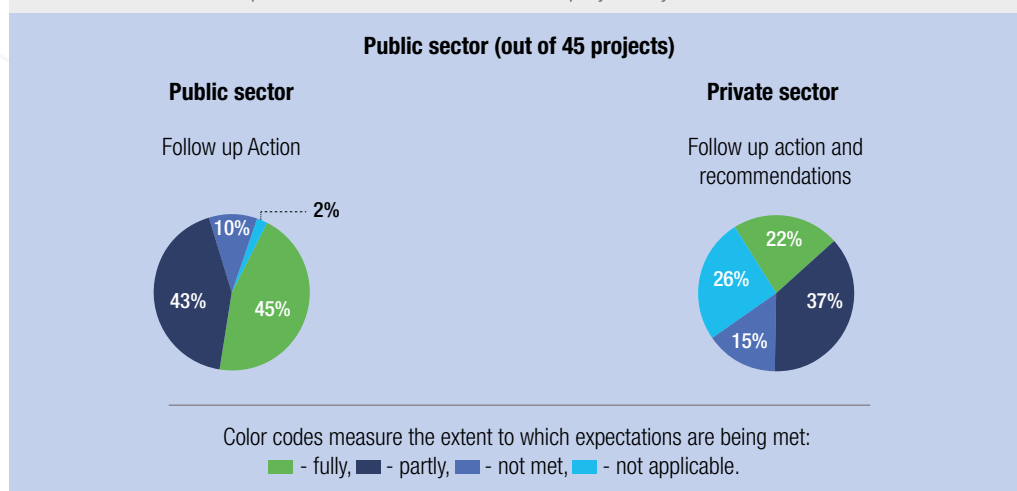
In this respect, the Delivery Dashboard does a better job than the IPRs at early warning and raising Management's attention. The portfolio dashboard inserts flags according to verifiable performance indicators and can categorize projects in potential or actual problem status even if the TMs think that all is well. IPR timeliness and validation are not captured by the Dashboard but it does allow an assessment of the reliability of the ratings. There are currently no results-related KPIs that could enrich the Delivery Dashboard.

MTRs are considered to be better vehicles for project adjustments. Performance upgrades did occur in five out of the 13 problematic projects that underwent an MTR in the QoS sample (45 projects). However, of the projects reviewed, only 40 percent showed evidence of any follow-up on the previous mission's recommendations for public sector projects and 27 percent in the case of NSOs (Box 7).

Non-Sovereign Operations

Recommendations of supervision missions are often not sufficiently actionable and vaguely described. The capacity to manage and mitigate risks is low, especially for LoC, since the Bank has limited influence over financial institutions to improve compliance and provide relevant data with respect to the achievement of DOs that could be used to trigger corrective action as needed. There is a tendency to use the ADOA results framework to forecast DOs years after a project has started, even when actual DO data should be available, and to leave the PSRs rating unchanged unless there is a major change. The expectation that the project development rationale and the logical frameworks in the PAR and the ADOA be aligned, could not be evidenced.

An NSO dashboard is available with the distribution of the active portfolio geographically, by financial instrument, by sector, and with KPIs such as the non-performing loans (NPL), and the weighted average risk ratings (WARR). The dashboard assesses performance through the value addition

Box 7: Minimal follow-up action on recommendations for project adjustments

to the Bank's income and return on investments, arrears and non-performing assets, the risk profile, and annual supervision status (based on PSRs undertaken during the year). It does serve as an alert system vis-à-vis the WARR, but does not cover other aspects such as for public sector projects.

The NSO project “watch list”³¹ is an additional tool to raise issues according to the status of the portfolio under: recovery/loss; rehabilitation; and active monitoring. The watch list is based on financial criteria with no assessment of risks to DOs. It is held under PINS coordination, which decides on the ratings and on whether the NSO should be under close monitoring or be managed by the Special Operations Unit (SOU) to find solutions. The SOU provides specialized knowledge and services to resolve issues of NPLs, including through restructuring, turnaround or in some cases recapitalization of distressed and underperforming companies.

Transaction costs of project restructuring

The nature of the problems to be addressed plays an important role for the incentive for TMs to raise and address issues, particularly if the project has to go through some form of restructuring involving

high transaction costs. TMs interviewed note that raising issues may cause additional work and most of them are already over-stretched. Formal restructuring is rare, both because the transaction cost is perceived to be too high and the Borrower is not always forthcoming, fearing to add delays. Interviewees noted that countries change priorities all the time but typically projects are not restructured to reflect the changes, and the Bank cannot proceed without the Borrower's agreement.

TMs interviewed conveyed a feeling that taking corrective action and project restructuring would be more proactively pursued if they felt that there was a support structure in place rather than being left entirely to themselves. TMs are typically very client-oriented and project-oriented, and are committed to pursuing project objectives, but the Bank needs to show that, in supporting the Borrower, there is an institution behind the TMs. The recent shift from project supervision to “implementation support”, the appointment of the Implementation Support Managers in each region and the stronger focus on decentralization go in this direction.

The majority of the 12 public sector projects reviewed for this evaluation showed the need for some degree of project restructuring. The main

reasons were initial procurement and disbursement delays, often linked to lack of readiness for implementation, changes in government priorities or local conditions, low capacity of the PIU, inadequate counterpart funds, complexity, and over-optimism. In most cases, targets were revised at the MTR (including major adjustments) but in no cases did the projects undergo a formal restructuring through the Board. The main reasons cited for this were the high transaction costs, avoiding delays, and the Borrower's reluctance. It was also argued that changing targets could be done without changing objectives and therefore no formal restructuring was needed. This creates a source of disagreement between Management and IDEV on the targets to use in validation. Contrary to Management's position, IDEV is required to use the formally approved targets in conducting the validation of the PCRs which, in the absence of formal restructuring, remain the appraisal targets.

This is also the case for CSPs/RISPs where there is a reluctance to go back to the Board at mid-term to modify coherence and relevance of the strategic pillars in an environment of changing priorities. As shown by the CSP case studies and interviews, the issue is now amplified with the required alignment to the High-5 priorities and possibly the need to retrofit existing strategies.

The restructuring of procedures is not sufficiently flexible and tailored to different degrees of changes needed, which would also allow realignment with IDEV's evaluation methodology applicable to revised targets in case Board approval is not needed. This mirrors similar concerns in comparator agencies³². If the Bank is adamant to achieve faster project design and approval, it should also recognize that, notwithstanding PD 2/2015³³, critical design activities (e.g., procurement plans, feasibility studies, validation of E&S Assessment studies [as reported in the ISS evaluation], setting up of PIUs, and baselines) are often rushed or carried over into implementation, which results in initial delays, greater inertia and more difficult problem resolution. TMs should feel more empowered to

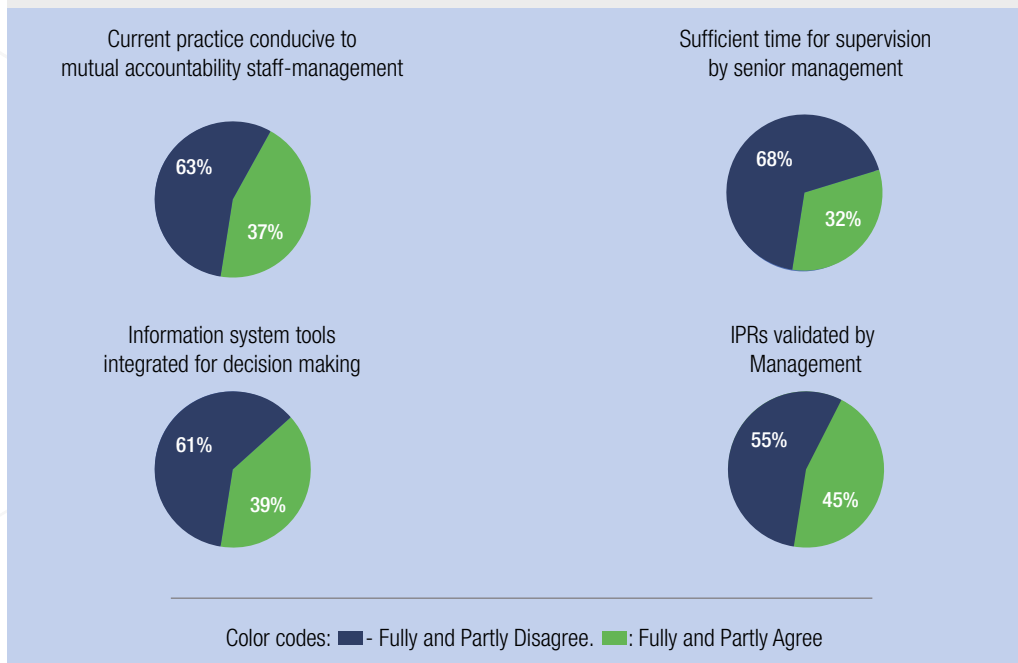
proactively raise and follow up on issues that can improve implementation performance, including the possibility of project restructuring, if necessary.

Promoting Accountability

One of the questions that the SESP evaluation tries to answer is whether the right degree of accountability is in place to allow the SESP architecture and its different instruments to function in a credible and effective way, signaling that the Bank is holding itself accountable for achieving results. Conversely, it is asking whether the SESP architecture is being implemented as a tool to enhance accountability towards Management and the Board.

Accountability is often the best way to achieve candid self-evaluation, meaning that someone in the hierarchy demands it. Procedures, guidance and rules are well in place and sufficiently clear. The main challenges are related to their implementation and the level of accountability that would be required for the system to work effectively. Staff survey responses from the QoS evaluation highlight a number of issues (Box 8). This requires looking at institutional incentive structures first. Management signals are very important in this respect, especially because accountability and incentives in the Bank have traditionally been skewed towards lending rather than the achievement of results. This is changing, however, and a number of measures have been put in place in the recent QA Implementation Action Plan (2019 - 2021) to establish a better balance.

Similar issues occur in all other comparator institutions, and this is understandable given that staff have greater control over project design and feel more accountable for it. Achieving results through project implementation and supervision is inevitably messy. It involves many stakeholders, especially governments, and there are many reasons for things to go wrong. Fixing issues is often beyond the capacity of staff, hence there is more reluctance to be held accountable for poorly performing projects.

Box 8: Accountability - QoS Survey Results (83 public projects) - Desk review

In looking at the two levels of accountability-internal and external-it is important to note that the two systems run in parallel and there is little involvement of operations in establishing the right links between the SESP outputs and the needs of the RMF. The 17 TMs and CPOs interviewed showed little awareness of what is required for the RMF and how the SESP can serve both purposes. Similarly, a number of RMF level 2 sectoral indicators (e.g., jobs created) are requested at the corporate level and trickled down to operations, even if the M&E systems may not be apt to provide the information required.

Internal accountability

Public Sector Projects

Self-evaluation by staff requires reliable evidence to function properly. Several factors influence the way accountability is being exerted in the implementation of the SESP:

- The effectiveness of M&E systems and how they link with result frameworks are at the core of credible reporting and accountability. As discussed previously, the results in this regard are modest, with issues raised both at the design and implementation stages.
- There is little reviewing and probing by Management of the ratings and underlying issues when clearing IPRs. Staff interviews confirmed that project supervision is not receiving adequate attention or follow-up from line managers. Validation of IPRs by Management is often done in a mechanistic way (when done), as managers have little time for IPRs in addition to the BTORs.
- Staff turnover is high as revealed by interviews and corroborated by the QoS evaluation, with insufficient handover. Three to four TMs are quite common during the life of a project. This results in a lack of continuity, reduces incentives for proactive action, and dilutes accountability.

Third-party validation is a major part of the accountability process, serving to keep the system honest. This can be done externally (IDEV) or internally (through peer reviewing or by a different division). The staff survey and the interviews confirm the usefulness of having independent ratings.

PCR quality is assessed by IDEV in the PCRENs. The PCR quality rating is affected by the same issue as project performance. Results are good when measured against the 2.5 rating benchmark (77 percent in aggregate over 2016 and 2017), as per the “range” methodology, but modest (52 percent) if measured against a 3 rating (Table 8). The PCRs’ quality situation can also be reflected through the average score, which remains at 2.67 and 2.9 in 2016 and 2017, respectively, revealing an important scope for improvement. However, a significant upgrading from the 2016 to the 2017 batch should be noted for the 3 and above ratings (from 30 to 64 percent), while timeliness grew worse (from 77 to 58 percent). Management’s own review of PCR timeliness in 2018 indicates a rebound to 85 percent as the result of a major push by the regional teams to improve timely completion. The majority of staff interviewed think that TMs should not do their own PCR. However, while using consultants under the supervision of TMs may provide some of arms-length review, candor does not improve much. Furthermore, consultants do not follow the same standards, which raises issues of comparability and quality. PCRs are typically given a lower priority by staff and there is little vetting of the choice of consultants.

The recent establishment of the Implementation Support Manager position could help address accountability issues by engaging on portfolio aspects with the sector divisions, providing the regional perspective and fostering better harmonization across sectors. It would move in the direction of a team approach (with the TM as the lead person), with the objective of improving candor and objectivity, and in line with the guidance of the Operational Manual. In this respect, CPOs would also have a stake and contribute towards a more collegial approach, to the extent that individual projects are also part of the country portfolio that the country/regional teams need to manage, hence benefiting from greater convergence.

Interviews carried out as part of this evaluation suggest, however, that in the context of increased decentralization, a better definition of roles is required and some concerns were raised as to the real capacity of this new function being carried out consistently on the whole portfolio, given the scarce availability of human resources. Generally, sector managers appear to retain a direct line of accountability over IPRs and ratings, but the practice seems to vary from sector to sector, and from region to region.

Even if some of the TMs interviewed have been prone to involve the Borrower in rating projects, this practice is not recommended, as it will reduce flexibility and managers will have even less of an incentive to question the ratings and exert accountability. Borrowers may, however, be part of the independent review process.

Table 8: Quality of PCRs

	2016 cohort	2017 cohort
# of PCRs rated 3 and above	15 (30%)	56 (64%)
# of PCRs rated 2.5 and above	40 (82%)	65 (74%)
Average PCR rating	2.67	2.9
Delivered on time	77%	58%

The Portfolio Dashboard (and quarterly portfolio flash report) is an accountability mechanism that ensures better compliance and provides a framework for enhanced proactivity on corrective action. Portfolio performance as illustrated in the Dashboard report is in the Performance Agreement of Vice-Presidents with the President. It could be cascaded down to managers. The concept of making performance of some key project indicators widely available is typically very powerful, especially if the system allows for the fixing of issues before they are brought to the attention of Senior Management. Making KPIs (disbursement, procurement, implementation delays, etc.) directly accessible by all staff is common in comparator institutions.

In the same vein, the rolling-out of the RRS would permit the posting, on a dedicated on-line portal, of key project implementation data and the most critical SESP products (IPRs, MTRs, PCR). It would potentially be a powerful instrument for improved accountability and reporting, and an effective tool for managers to streamline reporting requirements and easing the TMs' burden. Making information available widely is also likely to improve compliance by flagging omissions or delays. This addition is much needed and one of the areas where the Bank has been comparing poorly with comparator institutions.

Non-Sovereign Operations

The PAR results framework lacks precision and clarity in identifying development indicators and baselines. NSOs lack clear measuring tools of the progress towards DOs during implementation and data to report in PSR, ASR and XSR. Mostly, discrepancies between development results' indicators assessed at origination (PAR and ADOA) and those tracked during implementation, lead to poor monitoring and reporting of achieved results, and reduce the likelihood of effective risk mitigation.

An independent assessment of the quality of PSR, ASR and XSR undertaken for the purpose of this evaluation has revealed discrepancies in assessing DOs. There is a great tendency to rate project DOs as satisfactory despite evident shortcomings and data constraints. The E&S information is particularly scant (but there have been ad-hoc recent efforts to improve the E&S supervision of lines of credit). These issues reduce the effectiveness of the SESP in assessing and managing performances, particularly with respect to DOs and investment profitability. This is also due to insufficient knowledge of evaluation guidelines, and weak financial and economic analysis.

This notwithstanding, there is more of an internal validation system in place for NSOs than in the case of public sector projects. Accountability for portfolio management lies with the Credit and Risk Committee (CRC) and third-party portfolio monitoring is carried out by PINS, offering more of a dual accountability approach and oversight function (including over project rating). However, the low level of compliance seems to indicate low levels of accountability. DO ratings are not systematically validated by the CRC, nor is there a strong assessment of risks to DOs in the ASRs.

CSPs and RISPs

Country and Regional Programs (CSPs and RISPs) are all self-evaluated (by relevant country/regional offices) at mid-term and at completion (CSP-CR, RISP-CR). A CSP-CR is prepared within six months of the closing date of the CSP, and its findings should feed into the subsequent CSP. This means that the preparation of the CSP-CR is part of the preparation of the subsequent CSP. The CSP-CRs and RISP-CRs are prepared together with country/regional portfolio reviews (CPPRs) to assess the achievement of the strategic objectives and possible review of the results framework in line with the performance of the country or regional portfolios. CSP-CR and RISP-CR use no ratings but a notional assessment of satisfactory

or unsatisfactory and are sent to the Board (CODE) for information. Even if CSPs and RISPs are not rated by staff, they are both fed by the IPR ratings of the individual projects in their portfolio through CPPRs (since CPPRs and CSPs are presented in tandem), the reliability of which is uneven but typically rather optimistic. Therefore, the compliance and candor issues described above for projects permeate into CSPs and RISPs.

Sectoral, Thematic, Policy/Strategy Papers

There is no policy or official guidance on the application of the SESP to sectoral, thematic or policy/strategy documents (with the exclusion of CSPs/RISPs). The Operational Manual has a section on Economic and Sector Work that is still to be filled. However, concerned departments/units review their policies/strategies on an ad-hoc basis as part of the process for revising them or for preparing new ones. Some of these papers include statements mostly about MTRs, while CRs are very uncommon. When MTRs/CR are done, they are not rated. There is no systematic evaluation of MTRs/CRs provided by IDEV but, when it does, it reviews available self-evaluation reports. MTRs are presented to CODE or the Board for consideration.

External accountability

There is strong demand for accountability from the Board. The main reporting tool being used by the Bank to inform the Board is the RMF through the ADER report. The SESP is particularly relevant for the RMF level 3 indicators dealing with the Bank's operational effectiveness³⁴. Most of these indicators are fed by the SESP in an aggregated way. Therefore, the degree of accountability that can be exerted by the Board on operational effectiveness is only as good as the quality of the information provided through the SESP.

Interviews with ED's Offices raised a number of issues (see Annex 6) related to:

- The reliability of the RMF, the limited information received on the source of RMF data (and whether it is validated or not), and the role played by the SESP and how decisions are made.
- The low level of compliance with SESP requirements (IPRs, MTRs and PCRs) and their inconsistent implementation.
- The fact that the Board's comments are recurrent, but nobody seems accountable for addressing them.
- There is willingness to re-discuss the 4-point scale ratings methodology, particularly the adequacy and definition of the Satisfactory and Unsatisfactory ratings, and how they can be used consistently and effectively in projects, APPRs, CPPRs, CSPs and RISPs.
- The impression that the SESP is a box-ticking exercise. Staff turn to "satisfactory" very quickly, even when issues are present, and the narrative and the ratings do not match.
- The fact that CSPs are always "satisfactory". The methodology for CSP self-evaluation should change. The Board does not focus much on PCRs but mostly on CSPs. They are, however, concerned about the disconnect mainly between CSPs and CPPRs.
- The mixed quality of project-level logical frameworks, as they are not used consistently to engage with counterparts and to measure success. The main problem remains at measuring properly the outcome level.
- Conflicts of interest in having TMs preparing, supervising and then self-assessing their own work. This can only lead to candor issues.
- The need to eliminate the perception that project performance is associated with staff performance; build the reputation of "project fixers".

Public Sector Projects

A number of points can be made, as follows:

- In reporting the number of projects meeting their DOs, the RMF uses the PCR rating as a proxy, since the timeliness in the delivery of validated ratings for corporate reporting has been challenging. However, as noted in the Operational Manual and practiced by all other comparator organizations, the official rating should be the one assigned by IDEV. Furthermore, the PCRENs Synthesis Report (with contains IDEV validation ratings) is separately transmitted to the Board, often with different and lower ratings, which may be cause of confusion.
- A new target is more than 80 percent of projects being assessed positively by IDEV. However, results vary considerably depending on how “satisfactory” is defined. As discussed earlier in the report, using a rating range of 2.5 to 3.49 as the definition of “satisfactory” conveys a rosier image than if the 3-point rating was used as a benchmark, as described in the current 4-point rating system.
- Some RMF targets are set at a level which is not realistic given current performance or given the issues raised above. For instance, the target of at least 93 percent for sustainability seems achievable when using the PCR rating of 2.5 and above (87 percent for 2016 and 2017 combined), but is overambitious when using the PCREN rating of 3 and above (43 percent) or even 2.5 and above (71 percent).
- The “net disconnect”, which is the difference between the number of cases in which IDEV provides a higher rating (upgrade) and the number in which it gives a lower rating (downgrade), is a relevant indicator for the Board to consider. This indicator can be regarded as a proxy of the “candor gap”.

Non-Sovereign Operations

The performance of NSOs is not separately subject to external reporting through the RMF level-3 indicators, which may be related to higher degrees of confidentiality attached to NSOs. Contrary to other sister institutions such as AsDB and IFC, the Bank’s self-evaluation system does not inform the Board regarding NSOs’ performance at early maturity. The current practice does not offer the Board a timely and accurate independent opinion on achievement of NSO DOs and the Bank’s contribution and effectiveness. On a selective basis, however, projects are subjected to a comprehensive post evaluation using a cluster approach. This exercise is conducted independently by IDEV, and a Project Cluster Performance Evaluation Report is circulated to Management and the Board. Also, the annual Portfolio Status Report (based on PSRs) is sent for information to the President and the Board. As was previously noted, however, PSR compliance is low.

CSPs/RISPs

The CSP Completion Report is distributed to the Board and presented to CODE for discussion. A CSP MTR is prepared at the mid-point of the implementation period to assess implementation progress. If there is no change in strategy, the MTR is discussed at CODE level. In the case of change in the initial strategy, the MTR must be presented to, and approved by, the Board.

CSP and RISP self-evaluation reports are not validated by IDEV, contrary to comparator organizations. However, IDEV has recently validated one CSP and one RISP on a pilot basis, and will assess the scope for expansion³⁵. IDEV does carry out independent country program evaluations (CSPEs) on the basis of potential need for evaluative information (e.g., revision of Policy or Strategy when one expires) that encompass one or more CSP or RISP periods. These are distributed separately to the Board prior to the discussion of the next CSP/RISP. For this, IDEV shifted from a

6-point to a 4-point rating system in 2017 (in alignment with the Bank scale) on a pilot basis for two years. From 2014 to 2019, 22 CSPEs were prepared and discussed with CODE. The CSP/RISP-CR's assessment and the independent evaluation of CSPs and RISPs are not harmonized, hampering a full comparative analysis of what has worked or not based on common evaluation criteria and rating assessment of performance. Although an attempt was made to align country strategies to the High-5 priorities, the assessment is not based on clear evaluation criteria such as relevance, efficiency, selectivity, effectiveness, sustainability or risks to DOs, using KPIs.

Issues of consistency between CSP and CPPR outcomes have emerged recently, and the relevance of addressing at the same time portfolio management issues and the country assistance strategy has been questioned. CPPRs are mostly informed by the active portfolio (IPR ratings), which is affected by compliance and candor issues, while CSPs are informed by their own results framework and narrative. Currently, the CSP-CR is not included in the Dashboard; thus, it is of little value to help inform the new CSP. In this regard, the whole CSP process should be looked at to strengthen the accountability system and learning from past experience, and also to prepare for the next CSP. Furthermore, with the recent introduction of the Diagnostic Notes for CSPs and RISPs, the articulation with the CPPRs should also be tackled. Management has embarked on a discussion with CODE on the required revision of the CSP/CPPR methodology, which is now under way³⁶.

Learning

A main focus of the SESP is to deal with products meant to promote learning. This is especially the case with completion reports for projects, CSPs, RISPs, thematic reports, and sector strategies. In the case of projects, completion reports are validated by IDEV. It is therefore legitimate to ask how much learning actually happens and

whether such investment is producing the intended benefits.

Staff survey respondents (see Annex 7) are broadly positive that SESP outputs are good vehicles for learning (59 percent). However, individual interviews show a great deal of skepticism across the board. PCRs/XSRs are reported to be potentially good tools for learning, especially in the same country for follow-up projects, CSPs and CPPRs—assuming they are produced in a timely fashion. They are admittedly less relevant for staff in other regions, as these staff feel that lessons are too generic and not actionable, or too specific and not replicable. It was reported that learning from the SESP is happening sporadically, as PCRs do not always allow to extract a credible story that can be used somewhere else. The fact that some issues are recurrent and can be found in almost every PCR (poor quality of M&E, weak cost-benefit analysis, disbursement delays, procurement issues, low capacity of PIUs), as documented in the case study and the PCR Synthesis Reports, is a sign that not much learning is being translated into action.

The IDEV's Evaluation Results Database (EVRD) apart, there is no single place where SESP information and products can be accessed. The SESP has not led to a solid repository of knowledge that is mined and shared regularly by staff. Learning opportunities could be enhanced if the information were accessible online through a common platform. This is in the making with the advent of the RRS, which will include data from all SESP products.

There is a need for more creativity in terms of dissemination and sharing lessons with other countries on the same sector or similar specific issues. According to the PCREN Manual, periodic half-day workshops are to be organized with relevant staff to enhance feedback on findings drawn from PCREN/XSREN. However, this rarely occurs. Learning events should be evidence-based and be seen as a safe space where people are

willing to learn from success, as well as from failure. If the PCR/XSR is seen as a tool to judge the TM, it will undermine candor in ratings and learning opportunities. The rating itself is actually seen as an obstacle to learning because it potentially makes the discussion unnecessarily contentious. Learning should focus on generating and sharing knowledge with a forward-looking objective of adding value, providing solutions and improving a course of action irrespective of how it may be assessed.

There is therefore a fine line to be walked between the search for accountability and learning. Strategy papers (country or thematic) are more conducive to learning, possibly because no ratings are involved. Disagreement over ratings between staff and IDEV can further undermine the incentive to promote learning. TM interviews highlight that IDEV validation is often seen as too rigid and unfair, too focused on the initial results framework (that staff find too cumbersome to update) and they disassociate themselves from the process and the information that goes with it.

The little communication that there is between the PCR/XSR team and the IDEV team is not conducive to aligning views and an improved understanding of the issues. Debating and understanding the reasons for divergent views could actually be a source of learning in itself and improve the quality of the self-assessment function. This notwithstanding, the format of the PCR/XSR could also be improved to enhance learning, so that it is not seen as a perfunctory box-checking exercise at the end of the cycle. The majority of the staff survey's respondents agree that the lack of trust in the SESP undermines its learning potential.

Two of the most difficult aspects of evaluation and learning are the quality of the evidence being produced and the extent to which outcomes can be attributed to the Bank. The need for evidence-based lessons is normally addressed through the articulation between baselines, results

frameworks and M&E systems. In the absence of solid evidence, projects are penalized at the time of validation and learning becomes uncertain. Establishing attribution is harder and requires putting in place expensive statistically-proven methodologies, at least for a sample of projects. Occasional impact evaluation studies could be used on an ad-hoc basis to provide a solid set of evidence around which to organize learning events. Integrating impact evaluation methods into project monitoring systems could effectively provide quantitative and qualitative approaches to reporting, not only on the achievement of results but also on implementation issues.

Linking learning to the project cycle assumes that learning from PCRs can only happen once the project has closed, which means on average seven to nine years after the project started. PCRs are supposed to be delivered six months after closing and many of them (57 percent) incur delays and come too late even to inform follow-up operations, which are prepared before project closing. More flexibility and real-time learning would be required, which also means finding better ways to extract knowledge gained during supervision.

The lending culture has not helped foster learning through the self-evaluation processes. However, the signals are changing and a number of measures put forward in the QA Implementation Plan go in the right direction. Ultimately, the question is whether the same instrument can pursue two outcomes at the same time: accountability and learning, or whether the trade-offs are such that dedicated and distinct mechanisms and products are required. The WBG came to the same conclusion in its recent SESP evaluation report³⁷. While highlighting an excessive focus on ratings, it recommended that in looking for a better balance between performance, accountability and learning, the accountability function of the SESP should not be sacrificed and voluntary impact evaluations should be expanded to cover a wider spectrum of interventions in a given country, sector or region.

In the case of NSOs, the ratings and lessons learned are checked by sector managers in departmental meetings during which investment and portfolio officers normally attend to enhance the feedback loop. However, apart from the EVRD, which stores from IDEV evaluations, there is no consolidated lessons-learned database accessible by NSO development and portfolio officers. Lessons learned should be institutionalized and documented to ensure a strong capitalization of lessons learned based on operations, risk management, Macro-economic Policy, Forecasting and Research (ECMR) and IDEV interaction.

Furthermore, there is need to increase learning from NSOs on E&S safeguards and climate change adaptation, and also on SDGs and the mainstreaming of safeguards by reaching out to clients and private stakeholders. As noted in the ISS independent evaluation, this may need to start at project identification and preparation so that resources are allocated for project supervision. A couple of factors limit the capacity of the NSOs' SESP to contribute to learning: (i) the low level of compliance and hence of reporting; (ii) the required level of confidentiality and hence lower disclosure; (iii) delays in XSR validation that may impact the perception of credibility of the SESP; and (iv) the lesser emphasis provided to assessing

the achievement of development results compared with financial performance.

SESP adequacy to specific cases

The Operational Manual and some of the SESP outputs would need to be upgraded to enhance their relevance for a number of priority themes and operational products, such as:

- Gender issues are not sufficiently covered in projects, CSPs or RISPs, due to a weak methodology in assessing results (lack of outcome indicators). Gender experts are not systematically involved during the project cycle, CSP/RISP mid-term reviews and completion reports. There is no clear guidance in place to report on gender results in the SESP through the M&E systems, as evidenced by low ratings in the 137 PCRENs reviewed. No dedicated budget is in place to cover specifically gender issues during preparation, implementation and completion. A categorization system of gender issues, based on gender profiles, would ease follow-up and supervision, and increase the credibility of the SESP for reporting on gender in projects, CSPs and RISPs.

- Climate change and E&S safeguards aspects could be better integrated in the SESP. This is confirmed by the ISS independent evaluation, which found a low compliance with the Environmental and Social Action Plan (ESAP) requirements. The review of the available supervision reports found partial and vague reporting focused on certain environmental/social mitigation measures, but not an overall picture of the total measures included in the E&S management documents of clients and borrowers. Furthermore, the articulation of the climate change requirements with other E&S due diligence conducted by the Bank has been challenging and the corporate indicators used by the Bank, for instance in the Annual Development Effectiveness Review (ADER), are focused on work done before Board approval.
- Program-based Operations (PBOs) are not well served by the current SESP formats and templates. The business model is very different, as most PBOs only have one or two disbursement tranches and, when the project goes to the Board, first tranche's conditions have already been met and the project is fully or partly disbursed shortly after. Many sections of the IPRs are not adequate, nor is the reporting frequency. Therefore, design issues take priority and different arrangements need to be made to account for policy dialogue beyond the loan closing date (which is typically very short), in order to be able to assess how policy changes have impacted sector or macro performance.
- TA projects are subject to completing reporting only to the extent that their funding level is at least 1 million Units of Account (UA). This leaves out the majority of TA projects. For those that are covered, the format of the IPR and PCR could be simplified and made more suitable to projects that are essentially procuring consultants for capacity building, institutional strengthening and studies.
- Fragile and conflict situations are reported on during implementation as if they were operating in a normal environment. They would merit more specific and adapted formats, areas of enquiry, budgets and frequencies of reporting. Impact of fragility on outcomes should be more easily discernible through the SESP, as this is likely to improve accountability, to the extent that many issues related to fragility are beyond the capacity of the TM to fix.



Conclusions and Recommendations

Conclusions

The Bank's SESP are, to a large extent, relevant and aligned with comparators' best practices. The existing guidance tools and procedures are clear and aligned with the Bank's key policies, strategies (High 5s, TYS) and new business model (DBDM). A theory of change is not explicitly available for the SESP but, in line with the practice at comparator organizations, the system has consistently pursued three main outcomes: promoting better performance, fostering accountability, and enhancing learning. The SESP are closely integrated with IDEV's independent validation function, in line with comparator institutions. However, while ensuring IDEV's independence, the search for better convergence between staff and IDEV on ratings, harmonization of rating methodologies, definitions, and consistency of what is reported to the Board needs to be improved.

Results from the SESP evaluation show that the main weakness of the SESP is not in the established procedures, standards and norms but in their applications, particularly with respect to the low level of compliance with established procedures, deficient candor and over optimism in assessing performance, and limited resources for M&E during supervision. This has created an underlying and pervasive lack of trust in the system and a perception that the system is adding little value. This evaluation finds that progress towards a culture of development effectiveness has been mixed and the potential of the SESP to make an impact on the three main outcomes of performance management, accountability and learning, remains unmet.

This does not detract from the fact that project teams and TMs are fundamentally motivated to help clients deliver results and, by and large, are committed to project supervision. However, the prevailing enforcement and incentive structure, and the degree of accountability that is exerted, are not conducive to using the SESP as a reliable tool for reporting and an effective instrument for managing performance and taking corrective action. As one interviewee put it, "if Management cares about the SESP, the staff will care too". More accountability by managers should be accompanied by an effort to better assist teams with capacity and resources, and a better incentive structure that rewards pro-activity and corrective action and makes it easier to restructure projects.

It could be argued that the assumptions and expectations underpinning the SESP in terms of achieving its stated outcomes are unrealistic and that typically performance management and accountability have overshadowed or even undermined the pursuit of learning objectives. This finding resonates with the situation of comparator institutions. Rather than trying to fit the same instrument to achieve multiple objectives, it may be more appropriate to develop a distinct approach towards learning that has the SESP as one of its inputs but relies on more specific and adapted venues and instruments.

Managing ratings will remain a controversial subject as long as it is viewed as a tool to pass judgment on staff performance. Whether the system is too heavy on ratings or not remains to be seen, but at a minimum there should be an attempt to mitigate the

perception that ratings are easy to manipulate, that IDEV validation is out of context and mostly biased in nature, that the disconnect with IDEV can be largely ignored, and that the rating methodology is too rigid and bureaucratic. The ratings methodology needs revision to make it more objective and more evidence-based in its application. The experience of comparator institutions has been drawn to propose improvements.

While revisiting some of the underlying assumptions, incentives, and metrics that are likely to improve SESP performance, it may also be the opportune time to review the various SESP instruments for better alignment and consistency. At the same time, a number of outputs could be better tailored to specific thematic aspects and country situations.

Reporting tools such as Dashboards and the RMF have proven to be powerful accountability mechanisms and should be enhanced to cover indicators of compliance, results-based KPIs, and provide more visibility on possible “disconnects” in the assignment of ratings between staff and IDEV. SESP tools are relatively interconnected and the whole chain is as strong as the weakest link, starting with M&E systems and followed by the IPRs, which feed directly into other products such as the PCRs, the CPPRs, the CSPs and RISPS, and all the way to the RMF. Providing real-time information through online platforms and compliance data to line managers will enhance accountability and transparency, and facilitate automatizing the reporting of results, the timely escalation of issues to Senior Management, the reliability of information, and the efficiency of reporting at the corporate level.

Management is aware of these issues and, as the result of previous evaluative work, including by Management itself, has adopted, in the course of 2019, an Integrated QA Action Plan for both the public sector and NSOs. In particular, Management is committed to transforming the approval culture

of the institution into one that incentivizes and focuses on results and development effectiveness. It also wants to reorient supervision away from a passive compliance approach to pro-active and continuous implementation support. A major effort will be conducted to train staff through an Operations Academy. The Action Plans set in place a conducive environment on which to articulate the more specific findings and recommendations related to the SESP that this evaluation is putting forward as described below.

Recommendations

From the findings above, and considering the Bank's Integrated QA Plan, Management should:

Recommendation 1: Review the AfDB's results measurement framework and evaluation frameworks across the project cycle to ensure (i) internal consistency throughout the AfDB's results measurement and reporting system from operation to corporate level, and (ii) that there is harmonization between sovereign (SO) and non-sovereign operations (NSO).

Recommendation 2: Review and revise, in close collaboration with IDEV, the PCR/XSR ratings methodology in use including the ratings scale and guidelines in order to improve the reliability of the ratings system for all operations, and to better align SO and NSO.

Recommendation 3: Enhance the accountability for SESP products by developing appropriate indicators and explicitly incorporating these indicators in the AfDB's top-level corporate KPIs, in VPS and Directors performance agreements, and in staff performance discussions. The focus should be on delivery, timeliness, quality, usefulness and proactivity; encouraging candor in flagging and addressing emerging issues.

Recommendation 4: In the context of the AfDB's rightsizing, **review the SO and NSO staffing levels for M&E, and quality assurance activities** in order to ensure that the self-evaluation function is appropriately staffed and resourced.

Recommendation 5: Enhance compliance with corporate self-evaluation reporting requirements by clarifying the roles, procedures, frequency and reporting requirements for supervision, implementation support, and completion. In addition,

work with IDEV on an appropriate and realistic timeline for timely PCR/XSR validation in order to feed the formal annual development effectiveness report on the AfDB-funded interventions.

Recommendation 6: Enhance learning through SESPs by improving (i) quality of and accessibility of SESP products; and (ii) task managers' understanding and acceptance of SESP as learning tools rather than just as accountability tools.



Annexes

The following annexes are available on the website: <https://idev.afdb.org/en/document/evaluation-bank%E2%80%99s-self-evaluation-systems-and-processes>

Annex 1: Terms of Reference

Annex 2: Evaluation Matrix

Annex 3: Evaluation Methods

Annex 4: Evaluation Coordination Group (ECG) Practice Note

Annex 5: Comparators Review

Annex 6: Interview Summary

Annex 7: Staff Survey

Annex 8: A Snapshot from the Bank's Self Evaluation of Country Strategy Papers and Regional Integration Strategy Papers

Endnotes

1. Evaluation of Quality Assurance across the Project Cycle (2012–2017), October 2018.
2. The five areas of reform are: (i) reinforcing the knowledge of operations staff; (ii) strengthening project preparation; (iii) refining the assurance framework for quality at entry; (iv) sharpened focus on delivery and results; and (v) improving planning, budgeting and information.
3. A more detailed review of comparators SESP is contained in a working paper and available on demand.
4. The current scale is: 4 is Highly Satisfactory (HS); 3 is Satisfactory (S); 2 is Unsatisfactory (U) and 1 is Highly Unsatisfactory (HU).
5. The WBG, IFC, and IDB use a 4-point scale for individual indicators and a 6-point scale at the Outcome aggregate level, with three ratings above the bar and three below the bar. IFAD uses a 6-point scale for all situations and AsDB uses a 4-point scale.
6. IDEV. Synthesis Report on the validation of the 2016 and 2017 PCRs (2019).
7. The “net disconnect” is the difference between the number of cases in which IDEV provides a higher rating (upgrade) and the number in which it gives a lower rating (downgrade).
8. In 2018, IDEV has validated 100 percent of the 2016 and 2017 PCRs albeit under increasing budget and time pressure causing delays in the availability of ratings for corporate reporting. However, IDEV has since 2018 been validating on annual basis a sample of 65 of available PCRs and 100 percent of XSRs. Delivering the validated data during the first quarter of the following year would allow for timely incorporation into the RMF/ADER reports and address the current issue of having to use PCR ratings as a proxy.
9. IDEV Integrated Safeguard System (ISS) – Technical Report on overall compliance of the African Development Bank Group operations with the Integrated Safeguards System across the project cycle – Report – September 2019.
10. Information Note to CODE on the Bank’s Group’s Assessment Methodology of CSP Performance in CSP Completion Reports. December 2018.
11. This measure would complement the ongoing revision being carried out by PINS.
12. The High 5s are: (i) light up and power Africa, (ii) feed Africa, (iii) industrialize Africa, (iv) integrate Africa, and (v) improve the quality of life for the people of Africa.
13. Evaluation of Quality Assurance across the project cycle of the AfDB (2012–17) (2018).
14. IDEV. Synthesis Report on the validation of the 2016 and 2017 Project Completion Reports (2019).
15. The benchmarking review is available in Annex 5 and in a working paper (on demand).
16. For public sector projects: PAR, IPR, MTR, CSP/CPPR, PCR, PCREN. For NSOs: PAR, ADOA, PSR, ASR, XSR.
17. Comparator agencies have confronted similar situations and while IPRs are generally considered the main reporting instruments, the use of BTORs is left to management discretion.
18. The detailed assessment is presented in a working paper available on demand.
19. IDEV Integrated Safeguard System (ISS) – Technical Report on overall compliance of the African Development Bank Group operations with the Integrated Safeguards System across the project cycle (June 2019).
20. IDEV. Do Lines of Credit Attain Their Development Objectives? An Evaluation Synthesis 2010-2017 (2018). The report states that “There are no reliable data on the impact of LoCs on the final beneficiaries in terms of improved profitability or employment levels. Similarly, there are no data on the impact of green LoCs on energy efficiency and energy consumption. This is because: (i) FIs typically do not collect data on the impacts of their loans; (ii) FIs do not have management information systems in place to perform such tasks; and (iii) the IFIs exercise poor oversight during LoC disbursement.
21. The scale is: 4-Highly Satisfactory (HS); 3-Satisfactory (S); 2-Unsatisfactory (U) and 1-Highly Unsatisfactory (HU).
22. The “net disconnect” is the difference between the number of cases in which IDEV provides a higher rating (upgrade) and the number in which it gives a lower rating (downgrade).
23. The WBG managed to substantially reduce the disconnect over the past 10 to 15 years from some 10 to 15 percent to 2 percent. This is the result of full coherence of methods and ratings, strong public visibility and IEG having an impact year after year. In IFAD the disconnect is calculated differently as the difference of the overall ratings between the PCRs and the PCRENS and is reported to be -0.3 which is not comparable and tends to underestimate the issue.
24. Compare for example the AfDB’s satisfactory vs. unsatisfactory with the WB’s substantial vs. moderate, AsDB’s satisfactory vs. less than satisfactory or IFC’s satisfactory vs. partly unsatisfactory.
25. IDEV. Strengthening Country Strategy and Program Evaluations (2017).

26. PCRs/XSRs delivered in a given year are expected to be validated in the subsequent year. However, the actual time lag between PCR/XSR delivery and validation can be much longer. The 2016/2017 PCRs validations were combined and finalized in 2019.
27. New dimensions are being included in the ASR template and assigned a rating scale for better reporting and assessment of implementation progress: (i) project compliance with all covenants, (ii) project systems and procedures (corporate governance, financial management, M&E functions), (iii) project execution and financing; and (iv) project business success (financial performance, profitability and overall contribution to an enabling environment).
28. Cost-efficiency analysis of the SESP was constrained by lack of uniform data at the central level. Information was mostly gathered during individual interviews. Contrary to comparators (IDB, WB, IFC), the Bank does not use differentiated budgeting norms according to special context, complexity, and status of the active portfolio.
29. The WBG has also highlighted weaknesses in M&E and low capacity both in-house and in client country data systems. The Results Measurement and Evidence Stream (RMES) was put in place to strengthen M&E skills and most "results staff" were absorbed into Global Practices in the center. Capacity building is also offered by the Regional Centers for Learning on Evaluation and Results (CLEAR) initiative. This could lead to demand for more specialized skills to invest in good M&E. A similar CLEAR initiative is also being implemented by IFAD.
30. New guidelines are under preparation with new templates and valuation process of Bank assets. A corporate portfolio repository system making PINS the one-stop-shop for data on projects from inception to independent evaluations (XSREns) is being introduced.
31. 2017 Annual Quality Dashboard report.
32. Projects are placed on the watch list for reasons including: non-payment of principal or interest when due; significant decline in collateral value; borrower reports a loss either in half or full year results and there is evidence the loss was not due to extraordinary circumstances; financial difficulties of the borrower; poor information disclosure; poor corporate governance issues relating to the competence of management; major management changes, especially of key decision-makers without evidence of an acceptable succession plan; negative market trends, government directives; legal suits or threats of bankruptcy by other creditors; deterioration in the economic environment in general or in the industry the company operates in.
33. In the WBG it was recommended by the SES evaluation to increase flexibility in project design that minimizes the need to amend legal agreements as well as through simplified Bank and client restructuring procedures. Course corrections should occur as frequently as needed, informed by relevant and timely monitoring data.
34. Presidential Directive N° 2/2015 applies to the implementation, design and cancellation of sovereign projects and seeks to achieve five objectives: (i) improve the operations' quality at entry by requiring that necessary preparatory work (safeguards, procurement, project management) be carried out in advance; (ii) strengthen the implementation and management of projects; (iii) improve institutional efficiency; (iv) reduce the time for project approval; and (v) enhance transparency and accountability.
35. Level-3 indicators are: (i) increase the development impact of operations, (ii) enhance the quality and speed of operations, (iii) ensure strong portfolio performance, and (iv) increase the quality of Bank's knowledge and advisory service. Each of these indicators relies on a number of sub-indicators.
36. The pilot exercise for the South Africa CSP brought out some key issues, such as the absence of detailed assessment criteria or guidance on the ratings for the different dimensions for the CSP completion self-evaluation exercise. This leads to varied quality and content of the reports. The quality of the CSP results-based framework and its use as a monitoring tool needs to be improved, notably the linkages of proposed inputs, outputs and outcomes.
37. Information Note to CODE on the Bank Group's Assessment Methodology of CSP Performance in CSP Completion Reports (12 December 2018).
38. IEG. Behind the mirror: a report on the Self-evaluation systems of the WBG (2015).





IDEV

Independent Development Evaluation
African Development Bank



About this Evaluation

This report summarizes the findings of an independent evaluation of the African Development Bank's Self-Evaluation Systems and Processes (SESP). The evaluation is timely as the Bank revisits its strategic directions and operational processes. Over the past couple of years, the Bank has initiated an in-depth diagnostic of its quality assurance processes. This diagnostic was complemented by a series of Independent Development Evaluation (IDEV) evaluations that aims to examine the relevance, efficiency and effectiveness of the Bank's quality assurance processes across the project cycle as a way of enhancing their contribution towards performance management, accountability and learning.

The SESP are commonly defined as the assessments made of projects, country/regional programs, and policy/thematic review by the entity engaged in the activity. While the SESP are carried out by staff/Management, IDEV complements the process through the validation of certain products such as completion reports and through independent evaluations of projects, Country Strategy Papers and Regional Integration Strategy Papers.

The evaluation covers the period 2013 - 2018, and the implementation of the SESP over the time span that goes from project approval to closing (exit). It covers both public sector and non-sovereign operations. In addition to the project- or operation-level analysis, case studies for a sample of Country Strategy Papers and Regional Integration Strategy Papers have also been carried out.

The evaluation found that the Bank's SESP have many positive features. They lay out strong standards and procedures underpinning their functioning, as well as a cogent articulation with the independent evaluation function carried out by IDEV. However, the main weaknesses of the SESP are in the application of the established procedures, standards and norms.



IDEV

Independent Development Evaluation
African Development Bank

African Development Bank Group
Avenue Joseph Anoma, 01 BP 1387, Abidjan 01, Côte d'Ivoire
Phone: +225 20 26 28 41
E-mail: idevhelpdesk@afdb.org

