



Independent Evaluation Office
of the International Monetary Fund

BACKGROUND PAPER



BP/22-02/11

The IMF and Capacity Development— Monitoring, Evaluation, and Effectiveness

Ruben Lamdany

IEO Background Paper
Independent Evaluation Office
of the International Monetary Fund

The IMF and Capacity Development—Monitoring, Evaluation, and Effectiveness

Prepared by Ruben Lamdany*

July 27, 2022

The views expressed in this Background Paper are those of the author and do not necessarily represent those of the IEO, the IMF, or IMF policy. Background Papers report analyses related to the work of the IEO and are published to elicit comments and to further debate.

* Consultant, Independent Evaluation Office of the IMF.

Contents	Page
Abbreviations _____	v
Executive Summary _____	vii
I. Introduction _____	1
II. Findings of Previous Reviews of IMF CD and the M&E Framework _____	5
III. The Fund’s Framework for Measuring Effectiveness _____	8
A. Results-Based Management _____	9
B. Ex Post Evaluations _____	17
C. Five-Yearly Strategic Review of CD _____	26
D. Evaluation of Training _____	28
E. An Assessment of the M&E System as a Whole _____	29
IV. Evidence on the Effectiveness and Impact of IMF CD _____	30
A. What do RBM Data Tell About the Effectiveness of IMF CD? _____	32
B. What did Evaluations Find About the Effectiveness and Impact of CD? _____	38
C. What did the IEO Country Cases Find About the Effectiveness and Impact of CD? _____	45
D. The Findings from Training Assessments _____	50
V. What Are the Main Drivers of Effectiveness? _____	53
VI. Conclusions and Lessons _____	58
 Boxes	
1. RBM—Chronology of Fund Decisions and Implementation _____	10
2. RBM for CD in Other IFIs _____	11
 Figures	
1. Number of CD Internal and External Evaluations CY2012–2021 _____	26
2. Objective and Outcome Ratings _____	34
3. RBM Average Objective and Outcome Ratings Per Project for Completed TA _____	35
4. Average Objective, Outcome, and Milestone Rating by Project Status _____	36
5. Average Overall Satisfaction Ratings by Year, CY2010–2020 _____	50
6. Level 2 Indicators of CD Training by Year, CY2015–2020 _____	51
 Tables	
1. Projects Started Between 2013 and 2020 by Implementation Status as of End-2020 _____	14
2. Completed Projects: Rating Compliance _____	15
3. Definitions of the OECD-DAC Criteria _____	18
4. Completed Projects—Average Ratings _____	35
5. Rate of Satisfactory Completed Projects _____	35
6. Average Ratings of Objectives, Outcomes, and Milestones _____	36

Annexes	
I. Examples of Variables in a Logframe Template	62
II. RBM Information on Implementation, Compliance, and Effectiveness	63
III. Completed Project Assessments by Year and CDD, 2012–2021	68
IV. Internal Evaluations Undertaken by CDDs During 2011–2020	70
V. External Evaluations of IMF or RTAC CD, 2012–2021	71
VI. What Factors Correlate with CD Outcome Ratings?	72
References	87

ABBREVIATIONS

AD	Area Department
ADB	Asian Development Bank
AFE	AFRITAC East
AFR	African Department (IMF)
AFRITAC	African Regional Technical Assistance Center
AFW2	AFRITAC West 2
AML/CFT	Anti-Money Laundering/Combating the Financing of Terrorism
APD	Asia and Pacific Department (IMF)
ATI	Africa Training Institute
CAPTAC-DR	Central America, Panama, and Dominican Republic Regional Technical Assistance Center
CARTAC	Caribbean Regional Technical Assistance Center
CCB	Committee on Capacity Building
CD	Capacity Development
CDD	Capacity Development Department
CDMAP	Capacity Development Management and Administration Program
CEF	Common Evaluation Framework
CSN	Country Strategy Note
DAC	Development Assistance Committee
ESC	Evaluation Subcommittee
EUR	European Department (IMF)
FAD	Fiscal Affairs Department (IMF)
FARI	Fiscal Analysis of Resource Industries
FATF	Financial Action Task Force
FGN	Federal Government of Nigeria
FPAR	Final Project Assessment Reports
FPP	Financial Programming and Policies
ICD	Institute for Capacity Development (IMF)
IDB	Inter-American Development Bank
IFC	International Finance Corporation (World Bank Group)
IFI	International Financial Institution
LEG	Legal Department (IMF)
LTX	Long-Term Expert
MCD	Middle East and Central Asia Department (IMF)
MCM	Monetary and Capital Markets Department (IMF)
MDB	Multilateral Development Bank
METAC	Middle East Regional Technical Assistance Center
MOOC	Massive Open Online Course
M&E	Monitoring and Evaluation
OECD	Organization for Economic Co-operation and Development

PFM	Public Financial Management
PFTAC	Pacific Financial Regional Technical Assistance Center
RA	Resident Advisor
RAP	Resource Allocation Plan
RBM	Results-Based Management
RCDC	Regional Capacity Development Center
RCT	Randomized Controlled Trials
RSN	Regional Strategy Note
RTAC	Regional Technical Assistance Center
SCF	Somalia Country Fund
SPR	Strategy, Policy and Review Department (IMF)
SSA	Sub-Saharan Africa
STA	Statistics Department (IMF)
TA	Technical Assistance
TADAT	Tax Administration Diagnostic Assessment Tool
TAIMS	Technical Assistance Information and Management System
TSA	Treasury Single Account
TTF	Thematic Trust Fund

EXECUTIVE SUMMARY

This paper assesses the framework that the IMF has in place to monitor and evaluate capacity development (CD) activities, namely Technical Assistance (TA) and Training; it also presents evidence from different sources on the results of CD and the drivers of effectiveness. The main objective of IMF CD is to help member countries build strong institutions and boost skills to formulate and implement sound macroeconomic and financial policies. Current strategy emphasizes that CD should be country-centric, demand driven, and integrated with surveillance and lending. A robust monitoring and evaluation (M&E) framework is helpful to manage CD programs, and critical to assess the effectiveness of CD, distil lessons, and to report to the Board, Management, and donors.

The IMF has made gradual but significant progress over the past decade in strengthening the M&E framework for CD, but some key challenges remain. Among these challenges are how to make the use of results-based management (RBM) system a regular part of CD work, how to align M&E with CD's strategic goals, how to involve authorities more systematically in the M&E process, and how to ensure that systems, methods, and tools are user friendly. The paper examines the overall framework and the four main elements of the M&E framework: the RBM framework, the program of ex post evaluations, the five-yearly reviews of CD, and the evaluation of training.

Overall, the design of the M&E system is broadly adequate for the Fund's requirements, but it needs to be adjusted to reflect evolving CD strategic directions. The focus and coverage were appropriately enhanced to focus on results, to cover all CD independently of how it is financed, and to treat country-tailored training as TA. But the system still needs to be adjusted to take into account other strategic changes, e.g., there are no standards or processes to assess CD results in the context of the Fund's overall engagement with a member country nor to assess the synergies between CD and surveillance and lending. Also, the Fund needs to clarify how it will assess CD activities focused on global public goods and other Fund priorities that require a different evaluation perspective. The Fund also needs to raise authorities' involvement in monitoring and evaluation of CD and specify how findings and lessons should contribute to prioritization and to the design of activities.

The RBM system is already strengthening the results-orientation of project design and is providing data to report to the Board and donors, but challenges remain to its full implementation. The introduction of RBM has been slow as the IMF developed systems specifically for TA, had to acquire the IT infrastructure, and train staff. RBM was integrated into the design of the CD Management and Administration Program (CDMAP) and RBM functionality was operational as of the end of FY2022. To date, RBM has not played a significant role in evaluations and there is no clear guidance on how it should be used in CD allocation. Finally, staff becoming better acquainted with RBM processes within CDMAP may help reduce resistance, improve rating compliance and lead to greater use of the system's capabilities.

The IMF has introduced important improvements in how project assessments, self-evaluations by CD departments (CDDs) and external evaluations required by donors and partners are conducted, but there is a need for greater clarity on how these evaluations are to be used in the prioritization and design of future CD. In 2017, the Fund adopted a Common Evaluation Framework (CEF), updated in 2020, requiring a unified approach to external and internal evaluations based on the use of the OECD-DAC evaluation criteria and RBM log frames, and emphasizing learning over accountability.

- The self-evaluations undertaken by CDDs have improved over time in terms of their methods and presentation. They are designed as learning tools, some of them explore the determinants of successful implementation and outcomes, and the evaluation missions discuss with authorities their future CD needs. Given that very few CDD self-evaluations are produced, it is important that CDDs be very strategic in topic selection and that lessons be disseminated widely.
- Evaluation of regional centers and of externally funded programs are mandated by donors and conducted by external evaluators. Following the adoption of the CEF, external evaluations are becoming more standardized and easier to read. There is room to focus these evaluations more on issues of wider applicability, while still meeting donors' requirements.
- The assessments that are prepared at the completion of projects, until recently only for externally funded projects, are succinct, focused and with clear lessons. These assessments could be put to greater use as building blocks for higher level evaluations.

The five-yearly reviews of CD strategy have included elements of self-evaluation of Fund performance and facilitated incorporating lessons from experience into the process of adjusting policies and strategy. The most recent review, in 2018, conducted extensive consultations with recipients, partners and donors on the quality and utilization of CD products. This review provided the opportunity to examine the overall CD program with some time perspective and pointed at enduring weaknesses in the M&E system, such as the implementation of RBM and the need for uniform standards of evaluation. A strength of these reviews is that they allow evaluation to directly influence the policy development process, but there is a natural risk that evaluation findings for which there are no policy proposals be disregarded. Also, these reviews have focused on IMF policies and practices, but only marginally tried to assess the outcomes and impact of CD. From learning and accountability perspectives, policy reviews would benefit if they were informed by a comprehensive assessment of CD effectiveness.

The evaluation found that Fund CD has been effective in supporting member countries build stronger institutions and boost skills to formulate and implement policies across a wide range of issues and country circumstances. This overall assessment is based on the triangulation of evidence from RBM, evaluations, and case studies prepared by the IEO. This background paper presents evidence that provides a convincing picture that Fund CD has contributed to improving the technical capabilities of the recipient agencies, and it has helped strengthening the capacity of

governments and agencies to implement sound policies and improve outcomes. Country authorities and officials in implementing agencies appreciate the interactions with TA advisors for their high technical quality and relevant experience. They trust the IMF quality assurance and are willing to implement TA recommendations. Implementation generally led to the intended outcomes and, when sustained, to the broader desired impact. The country studies also illustrate how IMF training has built capacity across the world on many technical and policy issues.

Based on RBM data on project performance and effectiveness, the report concluded that more than 55 percent of the CD projects completed between 2015 and 2020 fully or largely achieved their objectives. The RBM database contains objective and/or outcome ratings for 137 of the 174 projects completed between 2015 and 2020.¹ Over the next few years, as the RBM database grows, it will be possible to analyze how the effectiveness of TA is evolving over time and to compare the outcomes across CDDs, beneficiary countries, and by mode of delivery. RBM data is already being used to explore the drivers of effectiveness; more data will enhance this work and make it more useful in prioritizing future activities.

Project assessments, external evaluations, and internal self-evaluations by CDDs also show a picture of significant effectiveness across a wide range of CD projects in countries with different income and capacity levels. Most of the reviewed projects had achieved or largely achieved their objectives, and some had a wider impact. Many CD programs in fragile states showed major achievements considering their context, although impact and sustainability were difficult to assess.

The expectations and assessment standards of the effectiveness of CD need to be calibrated further to the different country and agency circumstances. The country studies show that a strength of IMF CD is that the intensity and focus of activities are generally adapted to the different country and agency circumstances. Project design and log frames (i.e., milestones, outcomes, and objectives) are also adapted to some extent to the needs and circumstances of beneficiaries. However, to provide the basis for realistic assessments of results, it is important to further calibrate the expectations and assessment standards to more explicitly reflect the recipients' levels of human, institutional, and implementation capacity. The evaluation standards, including objectives and outcome targets, need to reflect the fact that what constitutes a partial achievement in one country, could be considered a great success in a different one. The standard for a fully achieved rating need to be set differently in a fragile state than in a well-functioning low-income country, and certainly should be completely different for a high-capacity emerging market economy. Further calibrating of RBM and evaluation standards to country circumstances would have led to higher overall ratings.

¹ RBM is designed to monitor outcomes and progress towards objectives through the end of projects, and therefore RBM data is not well suited for assessing impact. Still, by providing baseline information RBM can help prepare evaluations that sometimes assess projects' impact.

The evaluation found that Fund training led to significant learning gains, that participants utilize the skills and knowledge acquired, and that their supervisors and agencies consider that these skills have improved the working of their agencies. The Fund assesses most multi-country training using an appropriately modified version of the standard methods commonly used by trainers. The main evaluation methods are surveys of participants and their sponsoring agencies and employers, and pre and post training test results. Fund training is highly appreciated by participants, who show significant knowledge gain in post-course exams. As part of the integration of training with TA, there has been a move towards integrating tailored training into country and regional based TA programs, which should lead to more results-oriented training.

The evaluation evidence provides insights on the main drivers of CD effectiveness. These drivers can be helpful in the prioritization and design of the Fund's CD program. Key points are:

- The findings support the broadly held view that **country and agency ownership** is the main driver of effectiveness of CD projects and certainly of the sustainability of its results. Close cooperation with country officials in the design, implementation and evaluation of projects could help fostering ownership.
- **Integration of CD with lending programs and conditionality** increases effectiveness in the short run, but the evidence is mixed on whether outcomes can be sustained after the completion of the lending program.
- Countries and their agencies have a limited **capacity to absorb CD**, even when they are committed to its goals. Initial institutional and human capacity are key to the ability of countries to implement CD and achieve sustainable impact.
- The effectiveness of CD depends to a great extent on how it is delivered. Resident advisors (RAs) and long-term experts (LTXs) operating from regional centers enhance effectiveness by ensuring **better tailoring of programs to country circumstances, continuity, and support in implementation** (including after project completion).
- **Simple projects tailored to the specific country needs** enhance effectiveness, while complexity weakens it.
- Authorities' trust in the **technical quality of Fund advice and the relevant experience of CD experts** particularly because it is focused on issues where the Fund is seen as having world class expertise.
- **Cooperation and coordination** with CD recipients, partners, and within the Fund enhances effectiveness.
- **In-country training activities** delivered to intact teams and followed up by hands-on TA enhance learning and utilization.

I. INTRODUCTION

- 1. This paper examines the results of IMF capacity development (CD) activities, namely, technical assistance (TA) and training and assesses the systems that it has in place to monitor and evaluate these results.** This paper was prepared as background to IEO's 2022 evaluation of IMF CD and it focuses on the period 2011–2020. The paper examines the progress made in developing and implementing policies and practices in regard to the design, implementation, and utilization of the CD monitoring and evaluation (M&E) system and draws lessons for further improvement. It also presents evidence on CD results drawn from the Fund's M&E system as well as from country studies and the survey of authorities conducted for the IEO evaluation of CD. This introduction briefly discusses how CD activities have evolved at the Fund with particular emphasis on the implications for monitoring and evaluation. It also provides a roadmap for the remainder of this paper.
- 2. The main objective of IMF CD is to help member countries build strong institutions and boost skills to formulate and implement sound macroeconomic and financial policies** (IMF, 2019). The IMF's Articles of Agreement (Article V, Section 2(b)) state that upon a member's request the IMF can provide financial and technical services that would help promote the IMF goals of economic stability and economic growth. TA consists of activities aimed at developing institutional and policy making capacity, while training aims at building human capital among officials in member countries.
- 3. The purpose of the IMF's M&E system is to "review CD activities to ensure accountability; to assess their relevance, efficiency, effectiveness, and impact; and to distill lessons on how to strengthen future activities"** (IMF, 2019). Monitoring consists of collecting and analyzing information during project implementation on key indicators and comparing actual with expected results. Monitoring is designed to help manage the project and provide data to report on an ongoing basis to Management, the Board, and donors. Evaluation consists of the assessment of the results of the project relative to its intended objectives and outcomes. The Fund considers evaluation crucial "to foster learning from past experiences and enhance accountability" (IMF, 2017). The Board has repeatedly emphasized the importance of a robust M&E system to inform the development of strategy, the prioritization and allocation of resources, and the design and implementation of CD projects to best meet the needs of the Fund's membership (IMF, 2018b). M&E also can play a major role in gaining trust from the recipient authorities and securing support from donors.
- 4. The IMF's governance of M&E is similar to that of CD overall, but with Institute for Capacity Development (ICD) playing a more central role.** The Board provides strategic directions and oversight particularly in the context of budget discussions and periodic reviews. An inter-departmental Committee on Capacity Building (CCB) chaired by Management oversees CD planning, budgeting, monitoring and evaluation. The CCB sets evaluation priorities, approves the three-year evaluation plans, and reviews data from the Results-Based Management (RBM) system and evaluation findings. Since its creation in 2012, ICD has been responsible to develop

and monitor policies and procedures on monitoring and evaluation of CD Fund-wide and it is the secretariat for the CCB. This role has converted ICD into a champion for evaluation and has facilitated progress towards standardization of evaluation criteria and processes.

5. **Most TA is delivered by four Fund departments jointly referred to as Capacity Development Departments (CDDs):** Fiscal Affairs (FAD), Legal (LEG), Monetary and Capital Markets (MCM), and Statistics (STA), and by Regional Capacity Development Centers (RCDCs) run in partnership with donors and beneficiaries. Most multi-country training is delivered by ICD. Training activities that are tailored to specific country circumstances are designed, monitored, and evaluated as TA (these activities are referred to as tailored training).

6. **Over the past decade, TA has become more programmatic and focused on achieving results on the ground.** A typical TA activity used to consist of a country visit by a team of experts that would write a report on a specific institutional challenge or policy issue. The key goal was to deliver an excellent report with clear and relevant recommendations. The TA report would be submitted to the authorities and IMF staff would periodically inquire about its implementation as part of bilateral surveillance. By contrast, TA is now increasingly structured in multi-activity projects (aggregated into programs and workstreams) delivered over a couple of years; many of which are delivered by resident experts or experts located in the RCDCs. TA experts not only issue reports and recommendations but also work with officials to implement those recommendations to achieved project goals. As TA projects have become more complex, monitoring has become more elaborate, and evaluation now aims at assessing outcomes and impact rather than activities and outputs.

7. **Current CD strategy emphasizes that there should be a closer integration between CD and the priorities identified in Fund-supported programs and the policy dialogue in Fund surveillance—with implication for how CD should be evaluated.**¹ Area departments (ADs) are encouraged to prepare CD Country Strategy Notes (CSNs) that identify capacity gaps that hinder the ability of a country to implement agreed policies, list CD priorities, discuss how they relate to surveillance or program priorities, and reflect the authorities' views.^{2, 3} This strategy aims at ensuring that CD goals are more relevant at the country level and enhancing the effectiveness of activities. The integrated country-based CD strategy implies that the assessment of results should include how CD contributed to the effectiveness of surveillance and lending

¹ The 2018 Review (IMF, 2018b) re-emphasized the importance of country-centric, demand driven CD, replacing the prioritization system that had been put in place following the 2013 Review (IMF, 2013a), that placed greater weight on the Fund's strategic objectives.

² CSNs are mandatory for heavy users of CD. The CD strategy for other countries is less formal (IMF, 2020a).

³ Aligning CD with lending and surveillance has changed the roles of ADs and CDDs in setting priorities and implementation. While in the past, CDDs delivered TA in response to requests from recipients, ADs are now in the driver's seat in designing the CD strategy.

programs—in addition to assessing the outcomes of each CD project. It also has implications for how evaluation should be used in prioritizing and designing future CD.⁴

8. **Training has also undergone significant changes and is gradually being integrated into TA and its evaluation aims to be more results oriented.** The Fund, through the IMF Institute (that was merged into ICD), used to deliver training courses, mostly classroom based, on macro and financial issues. Some of these courses were cutting edge, others focused on the basics of policy making (e.g., the Financial Programming and Policies course, FPP), and in general most training activities were not country specific. Over the past decade there has been a move towards integrating training into country and regional based TA projects. Also, there is now more emphasis on evaluating the results of training through pre- and post-training testing of participants. Moreover, efforts are being made to measure how training impacts the effectiveness of the participants' agencies.

9. **Notwithstanding these important changes to the provision of CD, some key characteristics remain that are appreciated by recipients and that have implications for the evaluation framework.** First, the IMF continues to provide almost all CD free of charge to recipient countries, which generates significant excess demand. Second, IMF CD is designed and delivered directly by IMF staff or consultants managed directly by IMF staff or the regional centers, with continuous quality assurance by the IMF, which recipients view as a guarantee of high technical quality. Finally, a significant share of IMF CD is on issues that have global public goods characteristics (e.g., data dissemination, anti-money laundering). These characteristics of IMF CD preclude using costs and demand as measures of value.⁵ Under these circumstances, it is even more important for M&E to play a key role in informing the allocation of CD resources and the design of CD activities.

10. **Over the past decade the share of CD in IMF activities grew steadily, mainly supported by increased external funding, heightening the need to monitor, assess and report results.** The number of TA projects grew from 29 in 2014 to 70 in 2019, while the number of participant training weeks grew from 14 thousand to over 23 thousand during the same period. CD now accounts for about one-third of the IMF's operational budget, or about US\$400 million per year. The rapid growth in CD activities during the past decade was enabled by a large increase in contributions from partners and donors, now accounting for more than 60 percent of the funding. In addition to larger Trust Funds, donors and to some extent beneficiaries, helped to expand and create new regional TA and training centers. Partners and

⁴ Country ownership can “be enhanced by collaborative M&E efforts of past and current CD activities” and it is gauged in part by assessing implementation of past CD recommendations (IMF, 2019).

⁵ These modalities are different from those of TA from bilateral donors and multilateral development banks (MDBs), and this has implications for how different organizations monitor and evaluate their work. Bilateral donors often fund TA but subcontract its design, delivery, and evaluation to external consultants. TA supported by MDBs is often delivered within lending projects, where the borrower selects TA providers and pays for them out of the loan proceeds, giving an ex ante indication of the value for the recipient.

donors participate in the governance and management of these centers and of CD more generally, including on the M&E framework. Donors require evaluations of the programs they finance and many of them expected that these programs be designed, monitored, and assessed using an RBM system. Over the past decade, strengthening RBM and the evaluation program became key to address donors' reporting requirements.

11. **A robust M&E framework is helpful to manage CD programs and critical to understanding their effectiveness and impact.** Monitoring requires the systematic collection of data on the implementation of actions agreed with authorities and on progress towards intended outcomes. Regular monitoring of the performance of CD projects can help to make corrections in design and to adjust budgetary allocations during a projects' life; it is also key to be able to report to the Board, Management, and donors. The main goal of evaluation is to systematically assess the effectiveness and impact of CD projects and programs, and the workings and activities delivered by regional centers or within dedicated thematic trust funds (TTFs). Evaluations provide information that is useful in allocating resources, to improve the design of future interventions, and to ensure accountability to the membership and donors. Evaluations assess the achievements of CD projects relative to their intended objectives and to the baseline conditions, considering the costs and duration of a project. Evaluations are therefore facilitated by a well-functioning monitoring system that collects this information. The IMF is working towards implementing its RBM framework as such a monitoring system.

12. **The pandemic slowed down the implementation of M&E initiatives and the preparation of evaluations.** The Fund responded to the COVID pandemic by rapidly increasing virtual activities, for both TA and training, and by greater reliance on delivery by regional centers—an acceleration of ongoing changes.⁶ The Fund continued its efforts to implement RBM, albeit at a slower pace, because of the need to devote resources and energies to redesigning the delivery of CD activities. There were also delays in the preparation of planned evaluations as evaluations team had to shift to virtual interviews. While there have been significant adjustment difficulties the new modalities may have a positive impact on the quality and efficiency of monitoring and evaluation over the coming years.

13. **The analysis in this background paper is based on a combination of methods and sources,** including the country cases and thematic background papers prepared for the 2022 IEO evaluation, and IMF and external documents. We also reviewed the RBM system and data, and the information from surveys conducted as part of the IEO evaluation and staff's CD strategy reviews. Finally, we conducted semi-structured interviews with current and former country officials in recipient and donor countries, IMF staff, Regional Centers staff, and staff from other International Financial Institutions (IFIs) to learn their views on the Fund's monitoring and evaluation system and on the outcome of IMF CD. It is reassuring that the findings from different sources are consistent and mutually reinforcing and that prior assessments, including by IMF staff, had reached many of the same conclusions.

⁶ See Enoch (2022) for a discussion of the benefits and shortcomings of the evolving product mix.

14. **The remainder of the paper is organized as follows.** Section II presents the findings of previous reviews and evaluations of Fund CD related to M&E and the effectiveness of CD. Section III describes and assesses the systems in place to assess CD. Section IV examines the outcome of IMF CD and Section V discusses the drivers of CD effectiveness, i.e., the factors that affect the outcome of CD such as characteristics of the recipient and modalities of delivery. Section VI summarizes the main conclusions and draws lessons to improve the M&E system that could strengthen the results of IMF CD.

II. FINDINGS OF PREVIOUS REVIEWS OF IMF CD AND THE M&E FRAMEWORK

15. **The importance of strengthening the M&E framework for CD has been a recurrent message from IEO evaluations and staff CD reviews over the years and still remains relevant today.** This section describes the evolution of M&E policies and practices as seen by nine such evaluations, reviews, and policy documents issued during the past decade (plus the 2005 IEO evaluation of TA). The chronological narrative focuses on the challenges identified in these documents to strengthen the Fund's M&E framework—each of them showing that there had been gradual progress in the design and implementation of the M&E framework but pointing at important remaining gaps and weaknesses. The studies found that the introduction of RBM had been slow and often driven by the need to report to donors. The studies also concluded that evaluations had not been systematic, and they were not used for learning or programming. At the same time, these studies highlighted several positive features of IMF CD, among them its high technical quality and that it is greatly appreciated by authorities as trusted advice. But they found that CD had been more effective at skills transfer and providing advice than at institution building and that not enough was known about lasting results on the ground.

16. **In 2005 the IEO issued a comprehensive evaluation of IMF TA that found that the IMF did not have systems to track progress on major TA activities nor to identify reasons behind major shortfalls.** It found that the specification of outcomes and indicators that were monitored was often not clear enough to judge the extent of progress that had taken place. It suggested more candid ex post assessments reflecting authorities' views, greater dissemination of evaluation lessons, and increasing the synergy between TA and training. The evaluation found that TA had contributed to improving the technical capabilities of the agencies it supported, but that there was significant variability in the impact across projects.

17. **A 2011 Task Force on the Fund's TA strategy pointed at the challenges to implementing RBM and suggested that a more systematic approach should be used to evaluate the effectiveness of all Fund TA.** At that time, the Fund was trying to pilot RBM and was starting to develop logical frameworks to "capture and report on objectives and outcomes at country and topic levels," starting with externally funded projects. The Task Force already warned about many of the issues that still complicate the implementation of RBM, e.g., the long gestation of TA results that exacerbate problems with attribution and the lack of reliable outcomes indicators in many TA recipients. It also indicated that the volume of reporting and the mid-term and ex post evaluations of Regional Technical Assistance Centers (RTACs) and Trust

Funds needed to meet donors' requirements were growing due to the expansion of externally financed TA. The Task Force called for actions to reorient the evaluation systems and efforts towards assessing the effectiveness of TA and to develop a common framework to evaluate both donor-financed and internally financed TA.

18. **The 2013 review of the Fund's CD strategy (IMF, 2013a) proposed to develop a single strategy for TA and training, and to integrate CD with the Fund's other core responsibilities—surveillance and Fund-supported programs.** It found that the recommendations from previous assessments of CD had not been consistently implemented, underscoring the need for better CD governance. It highlighted that evaluation was crucial to foster learning and enhance accountability, and that it should provide feedback into prioritization and delivery of CD. The Board supported the plan to continue developing RBM to enhance planning and managing of outcomes and called to establish a common evaluation standard that would cover projects financed internally in addition to those financed by donors.

19. **The 2014 statement of "IMF Policies and Practices on Capacity Development" (IMF, 2014) stressed the importance of continuing the development of the RBM framework and to prepare a common evaluation framework.** The statement emphasized that CD should be demand driven and integrated with the policy dialogue that takes place under surveillance and Fund-supported programs. It highlighted the importance of country ownership to ensure effectiveness and impact, and that ownership could be enhanced by collaborative M&E of CD activities. The document indicated that the RBM framework and the planned common evaluation framework should apply to TA and training, and to all CD whether internally or externally financed. The statement also indicated that a periodic review would be submitted to the Board analyzing past performance in addition to proposing reform to make CD more effective.

20. **Two IEO studies, the update of the TA evaluation (IEO, 2014) and the assessment of self-evaluation at the IMF (IEO, 2015), found that the Fund had made progress upgrading the M&E systems for CD, but raised concern that these systems were focused on relevance and quality of inputs and outputs but not sufficiently on utilization and impact.** They found that monitoring of implementation (still using TA Information and Management System, TAIMS) was improving but reporting compliance was mixed with great variation across departments. They welcomed that work was ongoing to implement an enhanced Fund-wide RBM framework for TA and training but warned that initiatives had been complex, resource-intensive and had been driven mainly by donors' requirements. They found that CDD self-evaluations focused on learning, with some of them exploring the determinants of successful implementation and outcomes. On the other hand, the focus of most evaluations commissioned by donors was too diffuse as they covered many projects across many areas and countries. Like the 2014 statement, the IEO assessments recommended that the periodic reviews of TA and training should include self-evaluative elements on the relevance and effectiveness of past CD, in addition to proposals on how to move forward and that a process was needed to internalize evaluation lessons.

21. **In 2017, the Fund adopted a new Common Evaluation Framework (CEF) for all CD activities embracing the use of RBM log frames in project design and monitoring and applying the OECD-DAC criteria in a uniform way across all evaluations, including external evaluations.** The CEF was designed to streamline practices, increase comparability and use of results to foster learning and enhance accountability. The CEF codified good policies and practices that were already in use by some CDDs and brought the evaluation framework more in line with other IFIs. It established uniform methods and definitions, and required clarity about baselines, counterfactuals, objectives, outcome indicators and milestones. With regard to training, it emphasized the need to assess relevance, quality, and impact of courses.⁷

22. **The 2018 Review of the Fund's CD strategy (IMF, 2018b) highlighted the need to continue the progress in shifting CD focus on results on the ground rather than the Fund's own outputs, and in integrating CD with surveillance.** It recognized that building the RBM framework had accelerated since 2013 but remained "a work in process." It argued that for RBM to be an effective tool it would require much greater engagement with authorities on the objectives and milestones of projects. Key priorities in implementing the CEF were to develop a follow-up process including dissemination of evaluations and using them in the prioritization process. The review endorsed the implementation of the Capacity Development Management and Administration Program (CDMAP) that would link RBM to the management of budgets and personnel. The review emphasized that CD ought to be country-focused and integrated with surveillance—an implicit corollary of this strategic direction is the need to develop a framework to assess the impact of CD at the country level and the synergies between CD and surveillance.

23. **In 2019, the IMF issued an update to the Policies and Practices on CD statement, which emphasized integrating CD with surveillance and lending, country-tailoring, results-orientation and the need to continue developing RBM and the CEF.** Under this update, ADs in consultation with authorities would take the lead in determining country priorities that would be reflected in Regional Strategy Notes (RSNs) (and CSNs for heavy users of CD) outlining short- and medium-term CD priorities. The Strategy Notes would identify areas where CD could support the policy dialogue under surveillance and lending. These strategy notes, however, do not set a results framework for the overall Fund engagement in the corresponding region or country—a framework that would be needed to evaluate the effectiveness of CD if it were to be fully integrated with surveillance and lending. Once again, the statement explains that RBM is expected to play an increasingly large role in managing projects, as well as in monitoring and evaluation, and reiterates that departmental self-evaluations and donor mandated evaluations should apply the CEF, making evaluations more focused and easier to use.

⁷ Section III discusses the CEF and the Updated CEF in more detail.

24. **In 2020, ICD updated the CEF emphasizing learning over accountability, simplifying the approach, and aiming at greater utilization of evaluations.**⁸ The update aimed at addressing challenges that arose during the two years since the adoption of the CEF, e.g., the evaluation program had not been sufficiently strategic in the choice of topics, evaluations remained long and sometimes focused on CD content rather than on the achievement of CD objectives, and the reports were not widely disseminated.⁹ The Updated CEF advocated the use of RBM log frames as a starting point for evaluations and the consistent application of the OECD-DAC evaluation criteria.

25. **This chronology illustrates the persistent efforts the IMF has made over the past decade to strengthen the M&E framework for CD; but some key challenges still remain.** The large number of reviews and policy documents is an indication of the importance the IMF has given to improving CD and measuring its results. At the same time, as mentioned above, some themes keep recurring and remain current challenges. Key among these challenges is to make RBM and evaluations integral parts of the CD work, to align M&E systems with CD's strategic goals, including to support policy advice and program work, to involve authorities in the M&E process, and to ensure that systems, methods, and tools are helpful and user friendly. The next section examines how the Fund is dealing with these challenges.

III. THE FUND'S FRAMEWORK FOR MEASURING EFFECTIVENESS

26. **This section describes and assesses the IMF's framework to monitor and evaluate CD activities.** The focus is on the M&E framework, products, and activities, assessing, in turn, the four key elements of the IMF's multipronged M&E framework¹⁰:

- (i) the RBM framework (recently integrated into CDMAP)
- (ii) the program of ex post evaluations (project assessments, self-evaluations by CDDs and those required by donors);

⁸ Most of the CEF update report is structured as a detailed guide on the implementation of the updated framework, including how to apply the OECD-DAC and examples of relevant evaluation questions. This is an excellent innovation that should help staff (that are not evaluation experts) prepare more insightful and useful evaluations. In addition, this guidance will likely lead to more comparable evaluations that will allow drawing lessons of general applicability (although this was not the intention as the update downplays the need for evaluations to be comparable).

⁹ The Update also mentions as a shortcoming the fact that evaluations financed from the IMF's own resources are often led by staff from the department whose CD is the focus of the evaluation. IEO does not see this arrangement as a shortcoming, as discussed in Section III.

¹⁰ The IMF also seeks feedback on the relevance and quality of CD from beneficiaries and donors during the Spring and Annual Meetings, as part of its regular dialogue, and in the context of the steering committees of regional centers; it also conducts beneficiaries' surveys.

(iii) the five-yearly strategy review of CD (the latest being the 2018 Review of the Fund's Capacity Development Strategy (IMF, 2018); and

(iv) the evaluation of training.

27. **The Fund has another evaluation element, IEO's independent evaluations.** Several IEO evaluations focused on aspects of IMF CD. As mentioned in Section II, in 2005 the IEO conducted a comprehensive evaluation of TA that called for closer cooperation with authorities in the design and delivery of TA, it recommended greater focus on outcomes and to put in place a more robust monitoring framework. In 2014 the IEO prepared an update of the 2005 evaluation, and in 2015 the IEO examined self-evaluation of TA and training. Both these reports recognized the progress that had been made improving CD activities but emphasized that evaluations needed to assess effectiveness in addition to the quality of outputs. TA and training were also examined as part of IEO evaluations covering other issues, e.g., the evaluations of Research, Financial Surveillance, Fragile States, and Small States. These evaluations are not assessed in this paper (which is prepared as background to a forthcoming such evaluation).

A. Results-Based Management

28. **During the past two decades, at the request of its Board, partners and donors, the IMF has been gradually introducing a dedicated RBM system to plan and monitor the inputs, outputs, and outcomes of its TA.** The RBM system has multiple purposes. It aims to contribute to: (i) CD project design and management; (ii) strategic and resource allocation decision-making at the portfolio level; (iii) accountability and reporting; and (iv) evaluations (see IMF, 2021b).

29. Box 1 presents a brief description of the Fund decisions and efforts toward implementation of RBM.¹¹ As the chronology illustrates, it has been a long process. The Fund started to collect RBM data for some donor-funded projects in 2013, in 2017 it standardized the format and expanded coverage to include all CD projects, and in 2021 it became fully integrated into CDMAP.

30. The slow implementation is the consequence of challenges inherent in applying RBM to CD and some that are specific to the IMF. First, it took time to prepare the CD catalogue (objectives and outcomes that can be used consistently across the institution), relevant for the IMF context. Second, the IMF had to adopt and put into operation the necessary IT systems, and train staff. Third, it is difficult to change the mindset and working practices of a well-run organization, especially when it was good at what it had been doing (in this case, producing high quality TA reports). The 2018 Review argued that it takes at least four to five years to create an RBM system and then introducing an outcome focus in the actual work "requires changes in how

¹¹ See also History of Results Monitoring and RBM in the Fund, in Staff Background Studies of the 2018 Review (IMF, 2018c).

an organization is managed and delivers services, usually requiring behavioral and cultural change.”¹²

Box 1. RBM—Chronology of Fund Decisions and Implementation

- The 2002 Review of TA (IMF, 2002) recognized the importance of a greater results focus in line with international best practice and recommended that the Fund acquire a TA management system and systematize TA monitoring.
- Results-based monitoring was piloted from the mid-2000s at a project level for some externally funded programs, including selected RTACs and FAD’s Tax Policy and Administration and the Managing Natural Resource Wealth Topical Trust Funds. (2018 Review).
- In 2005, the Board endorsed the phasing in of the TA Information and Management System (TAIMS), a web-based repository for information and a monitoring tool for external TA projects.
- In reviewing the 2005 IEO evaluation of TA (IEO, 2005), the Board endorsed the recommendations to define ex ante results frameworks and to develop more systematic approaches to track progress of major TA activities.
- The 2008 review of TA (IMF, 2008) proposed to move toward “results-focused” project management. The Board stressed that project success should be assessed against benchmarks on outputs and outcomes, in line with international best practice.
- Following the recommendations of the 2011 Task Force on TA (IMF, 2011), a new RBM framework was established to: (i) monitor and evaluate the successes and failures of CD more effectively; and (ii) set priorities for resource allocation. Initially, RBM covered RTACs and TTFs only.
- In 2013, Staff started developing a standardized RBM framework, including catalogues and processes. The 2013 TA Review recommended the adoption of a comprehensive RBM framework for all CD and to systematize the focus on outcomes.
- CDPORT, an IT system geared to support operationalization of RBM and to modernize key aspects of CD management, was launched in 2015.
- In 2017, the RBM system became mandatory for all CD activities. Ongoing projects started since 2013 were retrofitted into the RBM system.
- The 2018 review of RBM by the Office of Internal Audit (OIA) (IMF, 2018a) identified that CDPORT was “not flexible enough to support a user-friendly interface” which was causing frustration among users and threatened user adoption. In response to these difficulties, Management decided to replace CDPORT by developing CDMAP, which aimed to promote a more user-friendly system with the potential to facilitate information sharing and coordination across the Fund.
- The 2018 CD Review explained that “sustained efforts will be needed to solidify a consistent and comprehensive results-based framework” and endorsed the implementation of CDMAP, a management and administration program that would capture and disseminate data on CD delivery and results (both financial and performance), to which RBM logical frameworks would be integrated for monitoring and assessing results. In the context of the 2018 CD Review, the Board once again indicated that it “looked forward to further steps to operationalize RBM.”
- July 2020 Release 1 of CDMAP incorporated the RBM catalogue and functionality for project planning.
- Release 3 of CDMAP took place in May 2021, transferring legacy RBM data from CDPORT. It launched functionality to prepare completion project assessments and risk management modules.

Source: IMF.

¹² It has also taken a long time for other IFIs to implement an RBM framework for their CD work, even those that had the advantage of already having functioning RBM frameworks for their lending activities (see Box 2).

Box 2. RBM for CD in Other IFIs

Other IFIs are also taking a long time to implement RBM frameworks for their TA and training activities. In most other IFIs the M&E framework for CD was adapted from the organization's framework for lending which is usually their main activity. This was not an option for the Fund since, in contrast with other IFIs, it conducts very limited evaluation work on its lending and surveillance activities. This gave the Fund the opportunity to design log frames and processes specifically for CD.

The activities and products considered TA and their modalities differ across IFIs, as do the monitoring and evaluation frameworks that apply to them. For example, many MDBs deliver much of their TA activities as components of their lending operations and are therefore financed out of the corresponding loans. In these cases, TA is assessed with regard to its contribution to the objectives and outcomes of the overall operation. Self-standing TA may or may not require a self-standing RBM log frame depending on whether it has its own objectives (e.g., developing capacity, generating knowledge, and providing policy advice) or whether it is part of a project preparation or to support its implementation.

The World Bank has an RBM framework that applies to its loans and grants. This framework is used to monitor and assess TA delivered as part of financing operations. The World Bank also prepares knowledge products, referred to as analytical and advisory services many of which would be considered TA products at the Fund; these products are assessed as part of the overall assistance to the corresponding country or as a specific theme in a thematic evaluation. Some analytical and advisory services are financed by donors that require the preparation of completion reports. At the IFC and the European Bank for Reconstruction and Development (EBRD), there are TA activities connected to lending and others aimed at advising governments on how to facilitate private investment, for which a completion report is prepared detailing the output and sometimes the expected outcome. In the IDB, some TA is part of lending projects. In addition, it provides self-standing grant financed (by IDB or donors) Technical Cooperation and a streamlined assessment report is prepared at completion.

Most MDBs continue to work towards improving their policies, systems, and guidance on how to apply RBM to CD, as they face some of the same difficulties faced by the Fund. Of note are the Asian Development Bank 2020 Guidelines for Preparing and Using a Design and Monitoring Framework for Sovereign Operations and Technical Assistance, and the 2020 Technical Assistance Completion Report Validation Guidelines that contain detailed instructions on how to build log frames and conduct evaluations of TA projects.

Source: IMF.

31. **The IMF has developed an RBM system—logical frameworks and processes—specifically for CD.** Logical frameworks or log frames are templates linking the different elements of a results chain—linking inputs, activities, outputs, milestones, outcomes, and objectives. Inputs, activities, and outputs are largely under Fund staff control—they include budgets, staff time, mission work, aid memoirs and TA reports and any other product or activity delivered to the recipient. Objectives and outcomes represent the benefits to the recipient and are pulled out of a catalogue prepared to provide some degree of standardization across projects.¹³ Milestones are indicators of the actions that the recipient needs to undertake to achieve the project benefits. For RBM purposes, the project design is required to clarify the causal links between the TA recommendations and milestones with the outcomes and objectives,

¹³ The catalogue of results was developed specifically for CD projects, it covers the main areas of IMF expertise, and is updated as needed. The RBM catalog includes about 100 objectives, almost 600 outcomes, and more than 1000 indicators. See Annex I for examples of objectives, outcomes, and milestones for each CDD taken from CDPORT.

as well as potential risks to the achievement of the project objectives.¹⁴ The Fund's RBM system was designed for and is fully devoted to CD, e.g., the processes and definitions in the RBM catalogue are aligned with typical CD objectives and outcomes. This is different than in most other IFIs where RBM templates were designed for lending and were adjusted to be used for CD. Using an RBM system that is tailor-made for CD should facilitate its implementation at the Fund, but this could complicate trying to use the RBM system and data to assess synergies between CD and lending.

32. **Tracking and rating milestones and outcomes provide a structure for management and monitoring of CD projects.** Each objective is linked to one or more measurable outcomes and observable milestones. Outcomes are the short- to medium-term results of the project (expected to be achieved when the recipient acts on the TA recommendations represented by the milestones). Progress is monitored by (preferably quantitative) indicators that are set at the start of the project. Milestones are usually actions to be undertaken or under the control of the beneficiary and are set to specific dates. Outcomes and milestones are supposed to be rated by project managers at least once a year in close consultation with authorities. Objectives are assigned a rating that in principle is a weighted average of the outcome ratings, but that can be modified by the project manager when warranted, e.g., if ex post it turns out that the mechanistic calculation does not reflect the reality on the ground because the link between outcomes and objectives were different than expected.

33. **Log frames are a helpful tool to design and manage CD activities and provide a template and incentives for greater results orientation.** Log frames provide a structure to help focus project design on its expected outcomes and objectives. In interviews, staff reported that defining the different components of the log frame enhanced the results orientation of project design. While in the past most of the attention of project managers was on the quality of the TA report, RBM and log frames increased the focus on implementation and results on the ground. Logframes are also a useful management tool, as rating milestones and outcomes help monitor progress toward the intended project outcomes and objectives. Objectives reflect the intended impact of the project, i.e., the high-level goals that the authorities and the Fund are hoping to achieve over the medium term. For each project, the log frame includes one or more objectives that are selected from a catalogue of objectives compiled for this purpose. Using standardized objectives and outcome indicators will facilitate comparing results across similar projects.¹⁵

34. **Authorities play a central role in guiding CD priorities—some role in selecting and designing CD activities (selecting objectives, and setting outcome indicators and milestones), but are much less involved in the development and monitoring of the RBM**

¹⁴ It contemplates five risk categories: authorities' political support, commitment by counterparts and their management, resource adequacy, external conditions, and other risks specific to each project.

¹⁵ For example, the FY2020 Annual Report of the Resource Mobilization Trust Fund uses RBM ratings to track progress over time and presents cross-country data on milestones and outcomes.

framework when rating them. CD guidelines call for involving authorities in the design of the project and in defining the log frame parameters, i.e., when selecting objectives and outcome, and setting milestones. Project managers also should then consult authorities when assessing ratings for milestones and outcomes and on implementation difficulties.¹⁶ There has been significant progress in involving authorities in the design and implementation of projects since the 2018 Review.¹⁷ In interviews, some staff and long-term experts (LTXs) (usually located in field offices or regional centers) explained that they set the milestones based on conversations with their counterparts but that outcome indicators are more technical. On the other hand, few of the interviewed authorities knew about the structure of the RBM framework or that a project had explicit indicators and milestones, and only one project counterpart remembered having been consulted about ratings. This is a missed opportunity because involving authorities in assessing progress can help foster ownership and effectiveness and is an important component of building capacity.¹⁸

35. **In countries with Fund-supported programs CD milestones are often set to parallel structural benchmarks.** While close alignment of CD with lending is an explicit strategic goal, it may have unintended consequences—authorities and CDD staff reported that aligning milestones with structural benchmarks sometimes leads to a less cooperative environment in the implementation of CD projects to the detriment of institutional building (see Legg and Sembene, 2022, and other country cases). Involving authorities in defining milestones may mitigate their concern that milestones may be used as proxies for structural benchmarks and lead to better outcomes.

36. **Almost all CD projects have now defined milestones and outcomes, a significant improvement since the 2018 Review.** The usefulness of RBM depends on staff complying with the requirements of defining the log frame parameters and rating them. As discussed in Section II, past assessments found that compliance had been uneven between externally and internally financed projects and across CDDs. As shown in Table 1, there has been great progress as almost 100 percent of the 432 completed or under implementation projects initiated during 2017–2019 had defined the log frame parameters.¹⁹ As of end-2020, RBM was tracking almost

¹⁶ The 2018 review urged that “Agreeing with the authorities on CD design and result frameworks upfront and using them as a communication tool throughout implementation should become standard practice.” And the 2019 Policies and Practices (IMF, 2019) indicated that “All M&E activities should seek inputs from recipient authorities to ensure that their views are fully taken into account.”

¹⁷ A majority of respondents to the IEO survey reported having been engaged to identify expected outcomes and milestones, as well as to assess implementation and effectiveness (see questions 20 and 21).

¹⁸ For example, Citrin and Legg (2022) indicate that authorities were not involved in the application of RBM and that past results were not taken into account in the annual TA planning process.

¹⁹ Annex II presents more detailed information on RBM implementation, compliance, and effectiveness. All the projects in the database had defined objectives and outcomes, and 95 percent had defined milestones. The projects missing information were Fund-financed, for which the use of RBM is more recent. This problem should disappear once RBM is fully integrated into CDMAP, as projects will not be able to proceed until all variables are defined.

400 CD projects under implementation and an additional 175 completed projects, with over 3,200 objectives and 6,450 outcomes. It also monitors the implementation of more than 17,000 milestones.

Start year	2013	2014	2015	2016	2017	2018	2019	2020	Total
Cancelled	0	0	0	2	3	3	2	0	10
Completed	11	26	10	25	68	29	6	0	175
Under implementation	0	3	15	26	174	85	62	27	392
Total	11	29	25	53	248	123	70	27	577
Percentage of projects with at least one defined milestone	82	83	92	84	98	99	97	89	95

Source: IMF, CDPORT database.
 Note: In 2017, the RBM system became mandatory for all CD activities. Ongoing projects started since 2013 were then retrofitted into the RBM system which explains the jump in the number of projects.

37. **Rating compliance has been uneven but is improving with the integration of RBM into CDMAP.**²⁰ Project managers are expected to rate at least once a year the implementation of milestones, the progress towards achieving the expected outcomes, and estimate the value of outcome indicators.²¹ Rating of milestones is a key monitoring exercise and can serve to engage authorities in a dialogue on how the implementation of agreed actions is progressing and help understand what is not working well. During 2015–2020, the Fund launched more than 500 projects with an average of 30 milestones per project. Overall, about 80 percent of these milestones were rated—a high proportion for a system that is still being phased-in. Still, the data shows some significant challenges, e.g., only in half of the completed projects were all milestones rated and in 10 percent of the projects no milestone was rated at all. Almost three-quarters of completed projects have all their outcomes rated (and a slightly higher share had at least one rated outcome). On the other hand, only 27 percent of completed projects had at least one of their objectives rated (and only 23 percent had all of them rated), which is puzzling given that objective ratings are, *prima facie*, calculated as a weighted average of outcome ratings (see Table 2 below). Notwithstanding the low rating compliance for objectives, these figures suggest that RBM processes are taking hold.²²

²⁰ Objectives, outcomes, and milestones are rated on a 4-point scale, where 1 stands for “Not achieved,” 2 for “Partially achieved,” 3 for “Largely achieved,” and 4 for “Fully achieved.” RBM ratings and project performance are discussed in Section IV.

²¹ Starting in 2022, project managers will be required to prepare an annual Project Assessment within CDMAP summarizing developments, ratings, and risks. The CDMAP functionality to support the preparation of the annual assessments should be available starting 2022.

²² Also surprising is that one-third of the completed projects (as defined by Bassanetti, 2021) were not flagged as such in the database. Most likely these are errors and lapses that will be corrected as the systems are fully functional.

Completion Year	2015	2016	2017	2018	2019	2020	Total
Outcomes	50	33	43	80	80	72	71
Objectives	50	33	36	32	22	16	23

Source: IMF, CDPORT database.
Note: Share of projects with completed projects that had all their objectives and outcome rated as of end-2020.

38. **In principle, RBM contributes to conducting ex post evaluations because it requires that project managers record at the start of a project baseline information and intended objectives and outcomes.** When creating a project in RBM, project managers have to document the initial conditions, objectives, outcomes, and outcomes indicators to measure progress and final results. RBM also ensures that data is being collected during implementation, and the objective and outcome ratings are a good starting point to evaluate projects. A revised RBM Governance Framework was rolled out in August 2020, adopting and standardizing good practices to support more consistent implementation of RBM.

39. **There is, however, some tension between RBM serving as a management tool and as the basis for evaluation.** As a project progresses and circumstances and CD delivery needs change, project managers can (and often do) redefine milestones and their expected timing as well as outcomes and their associated indicators. In rare circumstances they can also add or delete a project objective.²³ The ability to make these changes enhances the usefulness of RBM as a management tool but weakens its usefulness as a tool for ex post assessment.²⁴ Another limitation of RBM as a tool to evaluate the impact of CD is that impact (and sometimes outcomes) of CD often materialize with a long lag after the Fund has stopped tracking results, and that attribution of results is very difficult to ascertain.

40. **The RBM system is now embedded in CDMAP—an integrated operations management system that will help to align resources and administrative processes with intended results.** CDMAP brings together management of budgets, personnel, and results and provides a single platform for programming, managing, monitoring, control functions, reporting on results and costs, and reporting to donors, member countries, and the Board. Over time, if appropriately implemented, CDMAP will strengthen compliance with defining and rating log

²³ Changes in objectives, outcomes and milestones require consultation with stakeholders and in some cases approval from the CD portfolio manager and AD country lead and should be documented in the annual assessment of the corresponding project. It is expected that during 2022 and CDMAP Release 3, the initial parameters, including baseline indicators and original milestones, outcomes, and objectives will be saved in an audit trail. See Operational Guidelines – 2020 RBM Governance Framework (IMF, 2021b).

²⁴ It is expected that this tension will be mitigated following the implementation of CDMAP Release 3 as it will allow to save the initial parameters, including baseline indicators.

frame parameters, it will support the harmonization processes across departments, it will facilitate preparing real time reports on project implementation and outcomes, and it will enable to analyze efficiency (e.g., value for money). These are tall tasks that until now were (partially) being accomplished through several different systems (e.g., time and travel reporting, budgets, CDPORT). While from an RBM perspective there have been some transition pains in the move to CDMAP, it is clear that when fully functional it will strengthen the implementation and usefulness of RBM.

41. **Full buy-in of RBM by project managers is important to achieve its full potential benefits.** Managing CD projects with RBM has benefits, but it also has demands on project managers. Moreover, RBM is now embedded in CDMAP which is a new and complex system. Not surprisingly, the demands of RBM and CDMAP have elicited resistance from LTXs and other TA experts that see them as a tax on their time. Experience in other organizations and with earlier Fund initiatives suggest that if project managers feel that the system is a big burden on their time, they may not devote sufficient efforts to its implementation, which could lead to continued compliance problems or to RBM becoming a box-ticking exercise. There is some evidence of lack of ownership in this regard, e.g., some log frames are being copied from earlier projects without adjusting dates and descriptions. Low commitment to the system may also lead to creeping rating inflation over time. On the other hand, implementation of CDMAP and RBM is being overseen by an inter-departmental team of “CDMAP champions”—staff that understands and advocates for the benefits of RBM and CDMAP, and that helps project managers solve problems—a key success factor in implementing new initiatives.

42. **How are the RBM frameworks and data being used?** As mentioned above, the use of log frames is already strengthening the results-orientation of project design.²⁵ RBM data is commonly used by experts’ reporting on project implementation (e.g., in back-to-office reports) and to report to the Board and donors.²⁶ Completion project assessments prepared at the end of a project use the log frame structure and data as the starting point to assess results on the ground. On the other hand, the use of log frames and RBM data in CD evaluations has been limited up till now, in part because of a lack of data.²⁷ This should change as the availability and quality of RBM data increases and as evaluations follow the CEF. Also, IEO’s country cases and

²⁵ For example, already in its first year of RBM implementation, the introduction of milestones was mentioned as having strengthened the effectiveness of projects managed by CAPTAC-DR staff and their counterparts (see Lawson, 2018).

²⁶ ICD first used RBM data to prepare and illustrate its presentation to a Board Informal Session (To Brief) on CD Evaluations and Impact, in August 2021.

²⁷ IEO reviewed a sample of 10 external evaluations prepared during 2018–2020, of which only 2 used RBM data but without linking it to the corresponding log frames. Most of the others discussed the need to strengthen the RBM framework and data at the corresponding regional center or for the corresponding trust fund. IEO also reviewed a sample of four self-evaluations prepared by MCM and FAD during 2017–2019. While these self-evaluations focused on the OECD-DAC evaluation criteria and tried to assess results on the ground, they did not use RBM data, perhaps because it was not yet available.

interviews found that RBM information on track record played little or no role in CD allocations.²⁸ While in some cases projects were canceled or delayed because of limited traction, this was based on managers' and country teams' assessment and not based on RBM data. Finally, as discussed in Section VI, staff are starting to use RBM data to analyze the drivers of effectiveness of CD.

43. **In sum, RBM is already yielding some important results, but challenges remain to reaping the full benefits.** RBM has fostered a stronger results-orientation in CD, in the design, monitoring and completion assessment of projects. With the full integration into CDMAP it should contribute to better aligning resources and administrative processes with intended results. RBM provides information to report to donors on the results of the projects they financed, and it starting to be used to inform Management and the Board on the overall results of CD. On the other hand, authorities need to be more involved in defining, monitoring, and rating milestones and outcomes, baseline and implementation data needs to be saved to be used in evaluations, in prioritization and allocation and in project design, and there needs to be further efforts to reduce staff resistance.

B. Ex Post Evaluations

44. **Over the past decade, the Fund has strengthened the framework to conduct evaluations of its CD work.** The Fund has long recognized that "Evaluation and regular reviews contribute to good governance by fostering learning from past experiences and establishing a framework for accountability" (2013 CD Review), and conducted evaluations of its CD projects, but these were essentially ad hoc and limited in number. During the evaluation period, the Fund adopted a new evaluation framework, the CEF, and conducted different types of evaluations. Between 2017 and 2020, there were 61 project completion assessments, 11 evaluations required by donors and partners related to operation of RCDC and trust funds (i.e., external evaluations) and 6 self-evaluations by CDDs of particular CD topics (internal evaluations).²⁹ After a brief discussion of the CEF, each of these evaluation products is examined, highlighting the progress made during the past decade and the remaining challenges.

²⁸ There is no guidance on what role RBM should play in prioritization and allocation of CD resources. Of note in this regard is that effectiveness is not an explicit metric against which the CCB monitors delivery and decides on future allocations, which include "core topics, regions, country-groupings, growth areas."

²⁹ Between 2012 and 2016, there were 14 external and 16 internal evaluations.

The Common Evaluation Framework: Design and Implementation

45. **In 2017, the IMF launched the CEF (updated in 2020³⁰) that sets general guidelines and standardized the methodology for conducting evaluations.**³¹ Staff went about designing the CEF (both initially in 2017 and in the 2020 update)—by candidly examining the weaknesses of the different evaluation products and proposing measures to address them, with due consideration to resource constraints and the availability of information. The CEF aims at setting uniform methods for all evaluation products, centered around the use of RBM log frames and consistent application of the OECD-DAC evaluation criteria³² (see Table 3). It also sets the processes for prioritization, design and field work, and review and dissemination. The CEF emphasizes that evaluations should have a learning focus and that they should be disseminated widely. The goal is that application of the CEF, integration of RBM data and greater ownership of the evaluation process by IMF staff should lead to more relevant and useful evaluations for learning as well as for reporting to donors. Furthermore, it is expected that donor requested evaluations, CDD self-evaluations, as well as project evaluations would be prepared following the CEF.

Table 3. Definitions of the OECD-DAC Criteria

DAC Criteria	Key Evaluation Questions
<i>Relevance</i>	<ul style="list-style-type: none"> To what extent were the CD objectives derived from capacity gaps identified by the authorities, IMF surveillance/program, and other partners/institutions? Do the national authorities consider the CD objectives among the priorities of the country and/or agency? Was the design tailored to the country context?
<i>Coherence</i>	<ul style="list-style-type: none"> To what extent does the CD project support or undermine other IMF or partner interventions, or authorities' policies?
<i>Effectiveness</i>	<ul style="list-style-type: none"> To what extent were the CD outcomes and objective, as defined by the RBM framework, achieved or are likely to be achieved?
<i>Efficiency</i>	<ul style="list-style-type: none"> Were the CD inputs converted to outputs, outcomes, and impact in the most cost-effective way possible?
<i>Impact</i>	<ul style="list-style-type: none"> To what extent has the CD project enhanced the country's macroeconomic and financial stability and supported inclusive growth? What were the consequences on Fund strategic priorities, e.g., climate change and inequality?
<i>Sustainability</i>	<ul style="list-style-type: none"> The likelihood of the net benefits of the CD project continuing over the medium and long term, given the recipient country's capacity.

Source: Abridged version of Table A1 in 2020 Updated CEF.

³⁰ IMF (2020e).

³¹ This section focuses on the updated framework, referred hereafter as the CEF. The two main inter-related changes introduced by the Update in 2020 were a shift towards learning (de-emphasizing comparability and accountability) and eliminating the requirement to specify a full-fledged counterfactual when conducting an evaluation and instead requiring that project managers and evaluators collect and use baseline data.

³² See OECD (2019b) for the OECD-DAC Evaluation Criteria. Other references include OECD (2018; 2019a; 2019c).

46. **In addition to setting the framework, the CEF provides guidance on how to prepare evaluations that conform with the framework, which is particularly useful since most self-evaluations are prepared by CD experts rather than professional evaluators.** CEF guidance is quite detailed which is helpful, and it is sufficiently flexible not to discourage staff. The CEF provides guidance on how to design evaluations and on how to use RBM data and other information to strengthen evidence-based assessments. The CEF explains how to conduct the analysis using the OECD-DAC criteria, how to examine the relevance, coherence, effectiveness, efficiency, impact and sustainability of projects, thematic programs, or work done in a regional TA center. The aim of the CEF is to stimulate the design of shorter, more focused evaluation reports that will be more accessible and interesting to staff and authorities and would lead to better prioritization, and improvements in design and effectiveness of future CD projects.

47. **The design of the CEF seems appropriate, with a few caveats.** The CEF explicitly states that it applies to all evaluations included in the three-year work plan endorsed by the CCB, whether external, donor-mandated or CDD self-evaluations, which should lead to evaluation lessons of greater applicability across different activities.³³ It sets a process for the CCB to decide on the strategic three-year evaluation workplan, but it is less clear on how evaluation findings and lessons should be used in prioritization and design of CD projects and activities.³⁴ The design of the CEF is similar to, and in some ways more methodical than, the evaluation frameworks for CD in place in other IFIs.³⁵ It clearly sets the key principles, provides detailed guidance, and allows for adaptation to specific circumstances. The issuance of the 2020 CEF update shows that it can be changed as lessons from experience are absorbed.

48. **The evaluation of three high level CD goals would benefit from additional CEF guidance.** First, the CEF does not specify any product or process to evaluate the impact of IMF CD at the country level. Such evaluations would be more relevant in countries with significant Fund CD activity and could be based on an assessment of the impact of implementing CSNs. Second, the CEF does not provide guidance on how to assess the integration and synergies between CD on the one hand, and lending and surveillance on the other. Such evaluations would not be simple because (with very few exceptions) the Fund does not conduct evaluations of the effectiveness and impact of lending and surveillance. Third, there is no explicit discussion in the CEF on how to assess the relevance, effectiveness and impact of CD aimed at supporting global public goods.

³³ Staff is also encouraged to apply the CEF to evaluations that are not part of the CCB endorsed three-year strategic work plan. For example, the completion assessment reports discussed below are analyzed using the CEF methodology, even if it is unclear whether this is required.

³⁴ Adoption of the CEF in 2017 reenergized the practice of preparing a three-year rolling work program for TA evaluations on an annual basis that had been in place since 2003, but that had elapsed. The CEF instituted a more formal endorsement process of the program, that also includes the evaluation of training.

³⁵ One key difference between the CEF and the evaluation frameworks in most other IFIs, is that in those organizations CD is evaluated as part of, or in parallel with, the evaluation of lending activities.

49. **It is too early for a full assessment of the CEF's implementation, since it has only been in place in its current form since 2020, but it is already clear that there has been progress in many areas.** Every year, a strategic three-year rolling evaluation program has been presented and endorsed by the CCB that includes both externally funded and CDD self-evaluations. Most evaluation reports are shorter and more focused than in the past. Evaluations are now organized along the OECD criteria, and many of them provide ratings for each of these criteria (presumably allowing for comparability across projects, although comparability was eliminated as a CEF goal in 2020). Baseline information and RBM data are one source of information in the evaluations, even when information is limited at this time.

50. **However, implementation of the CEF has been uneven and there are significant gaps.** Some evaluation reports are still long and not reader friendly. Consultation with beneficiaries in the evaluation process is uneven, and most evaluation reports do not describe the authorities' perspective even it has been taken into account. Quality assurance of evaluations has been weak. Lessons and recommendations are not always couched in a way that would appeal to Fund staff beyond the project team, or to relevant audiences within the beneficiary country beyond the direct counterparts. Since evaluation reports are not widely circulated and read, it is difficult for their findings and recommendations to contribute to improve the design of future projects. Finally, each year ICD presents to the CCB a summary of findings and recommendations from past evaluations, but it is not clear how these findings and recommendations contributed to the prioritization of future TA program across countries, workstreams and projects, since expected effectiveness is only one criterion in the decision on how to allocate scarce resources (other criteria include demand, need, and Fund priorities).

Final Project Assessments Reports

51. **Project managers are required to prepare a completion report, shortly after the end of a project, assessing the project's achievements and shortcomings.**³⁶ Until recently, these completion reports were prepared only for externally funded projects, but have been required for all internally funded projects as well since 2021. In the future, as soon as the appropriate CDMAP functionality comes online, the assessments will be prepared within CDMAP whenever a project is closed, whether it was Fund or externally funded.

52. **Project assessment prepared upon completion of projects are succinct, focused, and discuss challenges and lessons.** The IEO reviewed a sample of 10 final reports of project assessments from FAD, LEG, MCM, and STA covering a wide variety of countries across area departments.³⁷ These reports were less than ten pages long (plus relevant tables), their analysis was conducted using the OECD-DAC evaluation criteria, and they presented the RBM log frames,

³⁶ For projects longer than a year, there is an interim assessment, though this does not apply to RTAC projects.

³⁷ These reports were prepared on a template generated in CDPORT, which serves as the basis for the CDMAP functionality. Annex III presents a list of completed Project Assessments prepared during 2017–2021. The Assessments reviewed are marked with an asterisk.

data and ratings (when available). Most assessment reports include a candid description of outputs and outcomes, and a couple of paragraphs on challenges and lessons. The reports mention interactions with partners and have a section to accommodate partners' statements (although there were no such statements in the reviewed sample).

53. **These end of project assessments could be put to greater use as building blocks for thematic and regional evaluations, and to distill lessons on how to improve future projects design.** The expansion in 2020 of the requirements to prepare final assessments to all projects is a welcome step. These assessments could become a key component of the evaluation system (as is the case in the MDBs where completion reports are the key building block for country and thematic evaluations). Preparing these reports has been and should continue to be quite inexpensive, particularly once the CDMAP functionality comes online, RBM is fully functional, and staff is more used to the OECD-DAC methodology. The reports should continue to be prepared by project managers shortly after project completion, when lessons and challenges are still fresh in their minds.

54. **Project assessments would become more credible and better learning tools if they included inputs from authorities.** The value of these assessments would be greatly enhanced if they included a brief statement from the beneficiary. While beneficiaries would not always be able to provide such a statement in a timely fashion, requesting their views would, at a minimum, lead to a conversation on achievements and shortcomings that could sharpen the reports' assessments. In any case, project assessments prepared shortly after project completion cannot cover the medium-term effectiveness and impact of CD—these are the goals of external and internal evaluations. Also, if they were posted on the intranet project assessments would become a key vehicle to disseminate lessons from experience across the IMF.

Self-Evaluations Undertaken by CDDs

55. **CDDs prepare self-evaluation reports with the goal of assessing the effectiveness of their projects and to extract lessons learned.** These evaluations are internal in that they are conducted by staff teams, sometimes supported by external consultants. These evaluations are designed as learning tools on specific aspects of the CDD work or on the impact of a long relationship with a CD user. Thematic self-evaluations examine the content, delivery mechanisms and traction of the CDD's projects. Some examine particular topics in groups of countries to assess the relevance of CD advice, its effectiveness and impact. CDDs also assess the effectiveness of groups of projects in intense CD users. Some evaluations explore the determinants of successful implementation and outcomes, and the evaluation missions discuss with authorities their future CD needs. The left bars in Figure 1 show the number of evaluations per CDD per year.³⁸ During 2012–2017, the Fund prepared 20 self-evaluations, an average of more than 3 evaluations per year. However, only three evaluations were issued since the

³⁸ Annex IV lists the 23 internal self-evaluations undertaken by CDDs during 2012–2021.

adoption of the CEF, two in 2019 and one in 2021. The steep reduction reflects the shift in resources to recast evaluations within the CEF and to integrate RBM data. In addition, staff explained that as a growing share of activities was externally financed it became more important to focus on improving external evaluations, a goal of the CEF.

56. **Self-evaluations undertaken by CDDs have improved over time in terms of their methods and presentation.** IEO reviewed eight self-evaluation reports, four from FAD, and two each from MCM and STA.³⁹ Most self-evaluations are based on the OECD-DAC criteria. Although most self-evaluations during the evaluation period covered older projects for which no RBM log frames had been developed at inception, they presented information on progress (or lack thereof) in outcome indicators, e.g., improved quality and publication of data in Guatemala (STA, 2015), changes in VAT compliance in several African countries (FAD, 2017), improvements in FOREX management in Malawi (MCM, 2017). In fact, as far back as 2012, the STA evaluation of TA and training for Peru, noted that “this is the first report that attempts to incorporate explicitly the OECD-DAC principles, reflecting on elements of the Fund’s emerging Results-Based Management (RBM) framework, which at the time of writing were in the initial stages of implementation.”⁴⁰ The evaluation of Debt Management TA in seven African countries (MCM, 2019) examined projects that did not have log frames and recommended the use of RBM in the design of follow-up projects. The evaluation of advice on Treasury Single Account (TSA) and Cash Management in selected Sub-Saharan African (SSA) countries went a step further and retrofitted milestone and outcome indicators to projects that did not have them at inception (FAD, 2019).⁴¹ Evaluations also reviewed the implementation of TA recommendations.

57. **In addition to examining outcomes, most evaluations assessed the quality of IMF work and some focused on the content of the advice, and sometimes reflected on the authorities’ views on these issues.** These evaluations are helpful to improve the design of future TA operations and help reassure recipient countries. The review of FAD TA across three East African countries assessed the ten-years’ experience providing TA on public financial management (PFM) and revenue administration to Kenya, Tanzania and Uganda (FAD, 2013). This report gave special prominence to the authorities’ feedback on the content, delivery, and results of the TA. The report on statistical TA and training to the Republic of North Macedonia

³⁹ The following internal evaluations were reviewed: Tax Policy and Administration Topical Trust Fund (FAD, 2013), Review of FAD TA across Three East African Countries (FAD, 2013), Evaluation of FAD Advice on Tax Policy and Revenue Administration in Africa (FAD, 2017), Evaluation of FAD advice on Treasury Account and Cash Management in Selected SSA countries (FAD, 2019), Malawi—Evaluation of MCM Multi-Topic TA on Central Banking (MCM, 2017), Africa Regional Debt Management TA Project Evaluation (MCM, 2019), Peru (STA, 2012), Guatemala (STA, 2015). See Annex IV.

⁴⁰ The evaluation explicitly states that it follows the recommendations of the 2005 IEO evaluation on Fund TA (IEO, 2005).

⁴¹ The evaluation report indicates that since “the Result Based Framework (RBM) was not in place during the time period of TA covered by the assessment, the team agreed with FAD a set of key performance milestones (see Box 1) to be used for measuring country progress in implementing CM reforms.”

(STA, 2013) discussed the modalities of delivery and the authorities' views on the impact of the TA and training. Some reports examine questions that are at the intersection between evaluation and research, such as the content of advice, or cross-country analysis of the link between TA and high-level macro variables (e.g., Sanjeev Gupta and others, 2013; and FAD, 2017 in Annex IV)—their analysis has an impact focus, but they are less able to establish attribution of results.

58. **Given that very few CDD self-evaluations are produced, it is important that CDDs be very strategic in topic selection, that evaluations be of high quality, and that lessons be disseminated widely.** Only three CDD self-evaluations have been produced since the issuance of the CEF in 2017. The size of the evaluation program is resource-constrained, but it is also determined by the absorptive capacity of the institution and by the parallel production of external evaluations (discussed below). The key elements to enhance the value of these few evaluations are to select topics that can enhance the effectiveness of future TA, more attention to quality control to ensure high quality reports, to make effective use of available RBM data, and greater efforts to disseminate lessons throughout the institution and to CD recipients. The value of evaluation reports as learning and accountability tools would be enhanced if they included a self-standing statement reflecting the authorities' assessments.

59. **The CCB supported by ICD could play a more active role in developing and implementing a strategy covering the prioritization, quality assurance and dissemination of evaluations.** The CCB is charged with reviewing annually institutional medium-term priorities for CD and to endorse the proposed evaluation programs, both internal and external. ICD, as the secretariat to the CCB, is well placed to help in preparing highly relevant evaluation programs, in consultation with TA recipients, donors, CDDs and ADs. Quality assurance of self-evaluation reports depends almost entirely on the evaluation team leader and her/his immediate supervisors since there is no structured inter-departmental (and in some cases no intra-departmental) review of self-evaluations. Inter-departmental review of CD evaluations would serve to disseminate their findings, in addition to improving their quality. In fact, dissemination of evaluation lessons is very limited outside of the originating units. This is a missed opportunity since some evaluation lessons have applicability beyond the corresponding unit and CDD. Current and former CDD staff reported that they conducted departmental seminars on their evaluation reports and that evaluation lessons impacted future design. It would be useful and not too costly, if CDDs would disseminate abstracts of evaluation summaries and lessons throughout the Fund and if they would organize seminars for wider audiences.

60. **The practice of having staff from the originating CDD lead the self-evaluations seems appropriate, given the role of such evaluations as a learning tool.** There have been recurrent discussions on having external consultants or staff from a different department lead the self-evaluations (most recently in IMF, 2020e). The idea behind this suggestion is that someone at arm's length from the corresponding CDD would be more candid in pointing at

shortcomings.⁴² This report takes the opposite view, i.e., that the people best positioned to conduct an insightful and useful self-evaluation are staff that had been involved in the projects being evaluated. Such staff are most aware of the challenges faced by the projects and would appreciate the opportunity to reflect on them and distill lessons on how to improve similar projects in the future. One obvious benefit from this arrangement is that lessons are organically fed into future operations, addressing the common difficulty of utilization of evaluations' lessons. The Fund can mitigate concerns about "beautification" of findings by the CDD by providing support on evaluation methodology and RBM log frames, either by ICD or external specialists, and by establishing a structured quality assurance process. But perhaps the most important factor in fostering candid self-evaluation is for the Fund to promote a learning culture where staff are not afraid of pointing at their own mistakes.

Donor-Required External Evaluations

61. **Most evaluations of RCDCs and of externally funded programs were required for reporting and accountability to donors and partners.** Most externally funded activities, which account for about 60 percent of Fund CD, are subject to mandatory periodic evaluations that, until the introduction of the CEF, were generally conducted following the processes, methods and presentation prescribed by these donors. For RCDCs and TTFs, these evaluations are usually conducted midway through each funding cycle and may contribute to the design of follow-up CD activities. These evaluations are a key accountability tool for donor countries and therefore until the introduction of the CEF, their timing, focus, methodology and structure were dictated by fiduciary requirements in those countries. These evaluations are conducted by consulting firms selected by the Evaluation Subcommittee (ESC) typically comprised of development partners, member countries, and Fund staff (IMF, 2020e).⁴³

62. **By and large, these evaluations served well their reporting and accountability goals.** Twenty-seven external evaluations were issued between 2012 and 2021, with significant variation from year to year and falling to only one per year since the beginning of the pandemic.⁴⁴ The right panel in Figure 1 above show the number of evaluations per year of RCDCs (a total of 13), TTFs (a total of 8) and bilateral accounts (a total of 5). While all of them served to report to donor governments, they differed in their specific goals. Some aimed to inform the decision process for the renewal of certain program (e.g., anti-money laundering/combating the financing of terrorism, AML/CFT, 2019; and Somalia, 2019), and others to justify the continuation of funding for a regional center (e.g., AFRITAC, 2018). Also, some external evaluations influenced the design of the corresponding RCs and TTF programs, particularly on governance issues.

⁴² External consultants and staff from other CDDs would most likely need to rely on the project team to learn about a project's shortcomings.

⁴³ The AD/thematic fund coordinator acts as the lead representative of the ESC and ICD acts as the secretariat of the ESC (IMF, 2020e).

⁴⁴ Annex V lists the external evaluations prepared during 2012–2021.

63. **Many external evaluations were very detailed and voluminous, seldom discussed actual results, and did not seek to draw lessons of general applicability.** To assess the quality and usefulness of external evaluations, IEO reviewed the latest 10 evaluation reports, 4 on RCDCs, 5 on TTFs and 1 on a country Fund.⁴⁵ Evaluations tended to include all thematic areas of delivery under the corresponding RCDC or TTF, and to examine governance and financial arrangements. There were significant differences in the quality of the analysis and in the presentation of the reports. Quality assurance of external evaluations is coordinated by the ESC. Relevant CDDs would comment on drafts of external evaluations, but generally these comments were prepared by an economist and not subjected to managerial scrutiny. Some evaluations drew lessons that were relevant but narrowly focused on the corresponding program or center, and as a consequence they were not of general applicability.

64. **Following the adoption of the CEF, most external evaluations are now conducted using the OECD-DAC evaluation criteria and RBM information but there is still room to enhance the readability and applicability of those evaluations.** While several evaluations refer to the CEF and RBM frameworks, the use of the methodology is uneven and they sometimes used different definitions of the relevant OECD-DAC criteria (e.g., CAPTAC-DR, 2018; and AFRITAC East, 2019). The 2018 external evaluation of the Debt Management Facility (DMF) assessed in eight countries if DMF was on track to achieve its objectives and goals, and to advise how to improve the program design and implementation. The report is very long and difficult to follow for those not directly involved with DMF.

65. **The differentiation between external and internal evaluations, beyond donors' requirements, is becoming less critical as both internal and external resources are deployed in support of most activities.** Both internal and external evaluations are currently structured around RBM and DAC-criteria. The key difference between these evaluations is that external evaluations need to fulfil the fiduciary requirements required by donors—and these requirements differ between donors. The total number of evaluations is constrained by the availability of resources for evaluations and by the capacity of the Fund and the TA recipients to absorb the lessons and recommendations of the evaluations. Planning both types of evaluations together would create more space for examining key issues that affect the Fund's overall CD portfolio.⁴⁶

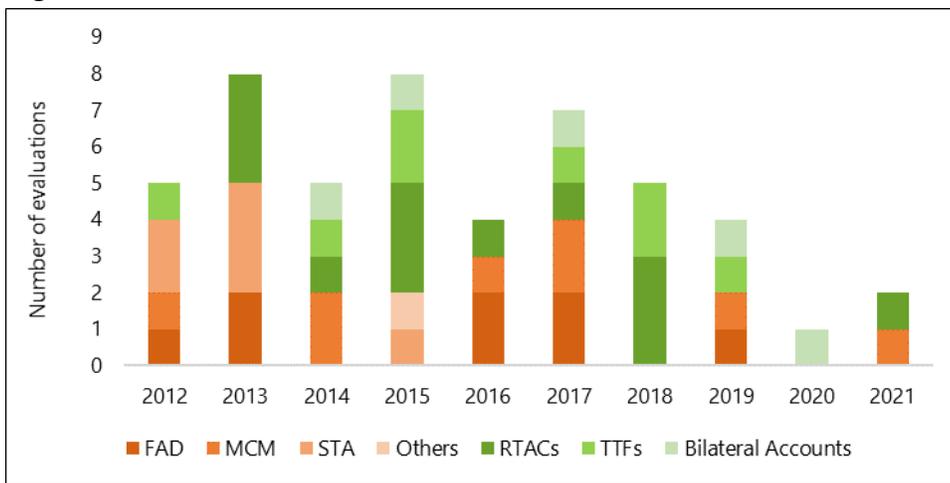
66. **ICD could coordinate the preparation of a strategic multi-year integrated evaluation plan, in close consultation with donors, RCDCs, CDDs, and ADs, for CCB approval.** Currently, ICD prepares a bottom-up three-years evaluation plan for CCB approval by aggregating the suggestions of CDDs and the requirements of donors. Instead, ICD could develop a strategic evaluation program that ensures that all critical issues are covered, in

⁴⁵ All the evaluations reviewed were prepared after the issuance of the CEF in 2017. See Annex V for a list of the evaluation reports since 2012.

⁴⁶ Figure 1 shows the yearly evolution of evaluations from different CDDs and external programs.

consultation with donors, RDCs, CDDs, and ADs. Such an approach would be facilitated by treating all evaluations as an integrated set of learning products and would enhance their relevance and quality as well as the utilization by recipient countries of TA, donors, and the Fund. Evaluations would still examine programs supported by donors, but these evaluations could be framed in a way that address questions that are of interest to wider audiences. In parallel, donors would likely have to conduct financial audits of the programs they support, but this would be less costly than the current arrangements. Evaluations could be conducted and perhaps led by IMF staff or consultants working on the project or facility being evaluated making them true learning instruments. ICD could provide quality assurance and, working with donors, it could disseminate the findings and conclusions of wider interest.

Figure 1. Number of CD Internal and External Evaluations, CY2012–2021



Source: IEO staff calculations.

Note: Internal Evaluations: FAD, MCM, STA, and others; External Evaluations: RTACs, TTFs, and bilateral accounts.

C. Five-Yearly Strategic Review of CD

67. **The FIVE-yearly reviews of CD strategy have included elements of self-evaluation and facilitate incorporating lessons from experience into the process of adjusting policies and strategy.** The Fund prepares a review of CD strategy every five years, in line with what is done for other policies and activities (e.g., Review of Conditionality, and Review of Surveillance). These reviews are conducted as the vehicle for policy development. Evaluation of Fund performance is, for the most part, an input to this process, and can draw on input from self-evaluation and external experts. The evaluative aspect of these reports focuses on Fund policies, strategies and activities—what the Fund did and how it did it. The system of embedding self-evaluations in the policy development process has served the Fund well in that it has led to better informed decision making.

68. **From an evaluative perspective, the system of five-yearly policy reviews has both strengths and weaknesses.** A key strength is that the review provides an opportunity for evaluation to directly influence policy frameworks and practices—this is most clearly reflected in the document on Policies and Practices that followed each CD review (e.g., IMF, 2014; 2019). Also, the periodicity of the reviews imposes a cyclical to revisit policies and practices that can be an antidote to institutional inertia, a risk in every large organization. But there are also risks to this process. The evaluative part of some reports is usually focused on issues for which there are policy reform proposals, which are not necessarily related to concerns on Fund performance or to results on the ground. Another risk is that in discussing future policies, Management and the Board may disregard the findings on past performance. Also, in line with other recent Fund policy reviews, the reviews of CD strategy do not examine in any depth the actual results on the ground from capacity development activities—were these activities effective; did they achieve their intended outcomes and impact? This gap is not necessarily a system flaw, as long as these questions are addressed elsewhere ahead of launching the policy review.

69. **The most recent strategy reviews of CD, in 2013 and 2018, discussed past Fund performance with an evaluative perspective.** The 2013 review took place shortly after the creation of ICD and focused on assessing the overall CD framework and governance. It pointed at enduring weaknesses in the system, including the slow and uneven implementation of RBM, and the wide differences in the focus and methods between external donor-required evaluations and internal self-evaluations, and between self-evaluations in different CDDs. Building on the findings of the 2013 review, in 2014 the Board endorsed the application of RBM to all CD, called for the development of a common methodological framework for evaluations, and required that recipient authorities' views be fully taken into account in M&E reports. The 2013 Review explicitly stated that future policy reviews should have a backward-looking component that should evaluate the implementation of the initiatives agreed in the review, including in regard to prioritization, funding arrangements, M&E, and delivery.

70. **The 2018 review was an example of good practice in terms of including assessment of past Fund performance.** The review conducted extensive consultations with recipients, partners, donors and staff on the quality and utilization of CD products. It conducted a survey of stakeholders, and included a series of background papers assessing, among other topics, progress in the areas indicated in the 2013 review, i.e., prioritization, funding arrangements, M&E, and delivery; it also included comments by external experts on key issues. On M&E, the review found that RBM was still a work in progress and that "effective use of RBM will also entail greater engagement with authorities and their buy-in on proposed CD project objectives and milestones." The review noted the "foundational work on the common evaluation framework" and that there was no clear governance for RBM and the CEF. It also found that dissemination of evaluations' lessons and reports needed strengthening and that evaluations had not played a major role in prioritizing across countries and issues. While there has been progress, some of these concerns remain relevant.

71. **Future CD strategy reviews should continue to include a backward-looking component that assesses of Fund performance, and they should be preceded by an evaluation of the effectiveness of past programs and projects, even on issues where no changes are being considered.** The 2013 review highlighted the importance of having a backward-looking assessment of Fund performance as part of policy reviews to serve as an input and provide credibility to the forward-looking policy reform proposals. The 2018 Review devoted significant resources to such an assessment and identified changes that could enhance the effectiveness of Fund CD. However, neither of these Reviews examined in depth the effectiveness of CD, nor were they based on a separate comprehensive assessment of what CD had achieved. From learning and accountability perspectives, policy reviews would benefit if they were informed by a comprehensive assessment of performance and actual outcomes and impact. The IEO evaluation will provide such assessment for the next CD review, scheduled for 2023. A more integrated and more strategic approach to CD evaluation, as recommended here, would provide the necessary inputs on performance and outcomes for reviews thereafter.

D. Evaluation of Training

72. **The Fund assesses multi-country training using standard evaluation methods.**⁴⁷ For most training courses the Fund uses an appropriately modified version of the four levels assessment method commonly used by trainers (Kirkpatrick, 1994):

- **Level 1 measures participant satisfaction.** This is measured with a survey conducted close to the end of each course. Participants are handed a questionnaire and are encouraged to provide suggestions on how to improve the course.
- **Level 2 assesses participant learning.** Learning is measured using pre- and post-training exams. While these exams are better suited to lower-level officials and IMF staff, ICD is committed to perform Level 2 evaluations across the board. ICD developed an index of learning results based on the increase in average test scores between the pre- and post-course tests and the share of participants achieving 60 percent and above on the post-course test.
- **Level 3 assesses the training outcomes for individuals and institutions.** Outcomes usually refer to changes in participants' effectiveness at their workplace as a result of training. ICD tries to assess effectiveness by using two sets of follow up surveys. These surveys are administered by external vendors to encourage respondents to be candid. ICD conducts occasional surveys of participants and their supervisors or sponsors several months after the conclusion of training events. Participants and their supervisors are asked similar questions about whether the training is relevant to their current work, whether it had an impact on work performance and on job responsibilities. Also, every

⁴⁷ A thorough description of the IMF training program and discussion of how it is evolving can be found in De Lannoy (2022).

three years, ICD conducts a broader survey of central banks, ministries of finance, and statistical and other government agencies in most member countries. The triennial survey objective is to gather sponsoring agencies' views of the effectiveness of the ICD training program and information about future training needs. Seven surveys have been conducted since 1995, the most recent one in 2019. Surveys ask about the use of knowledge and tools acquired through IMF training, and about factors that could improve training effectiveness.

- **Level 4 assesses impact in terms of improved performance of the participant's organization.** The Fund, like other IFIs that provide training, does not conduct Level-4 evaluations because they are too costly and difficult to implement.⁴⁸

73. **Overall, the Fund has made significant efforts to assess the effectiveness of training, a difficult undertaking that will become more viable as training is further integrated with TA.** It is particularly challenging to assess the effect of training on institutions, let alone countries, because it is an activity targeted at individuals. In fact, it is even difficult to assess the lasting effect of training on the individual participants. Like most training organizations, the IMF gathers information on the relevance, quality, and delivery of courses, as well as on the acquisition of knowledge and skills (Levels 1 and 2 above). In addition, the Fund conducts surveys to gather information on the effect the acquired knowledge and skills had on the participants' agency. The surveys conducted by the Fund (Level 3 above) are as good a measurement tool for outcomes as those used in other IFIs and training organizations. Tailored training is gradually being integrated into TA and country programs, and therefore planned and assessed using RBM frameworks. This will refocus their assessments onto outcomes and impact, but these will be embedded in those of the overall Fund intervention, thus diluting attribution. In any case, Level 1 and 2 assessments would still be necessary to learn about quality of content and delivery. Over time, the integration with TA should lead to more results-oriented training and it should allow to get a more systematic perspective on training effectiveness.

E. An Assessment of the M&E System as a Whole

74. **The Fund has now a wide-ranging multipronged M&E framework that could support learning and accountability and enhance the effectiveness of IMF CD.** The four elements examined above (i.e., the RBM framework, the program of ex post evaluations, the five-yearly review of CD, and the evaluation of training) could cover the full CD cycle from strategy and project design to implementation and distilling of lessons. This M&E framework is sufficiently flexible that it could be adapted to changes in CD strategy and scope. RBM provides a structure to help design results-oriented projects and to monitor progress during implementation, and to

⁴⁸ In the past, the IMF Institute conducted tracer studies to assess the impact of training, but they have appropriately been discontinued. These are studies that "trace" participants years later and assesses how their contributions to high level goals have increased. Tracer studies are very costly, and attribution of results is difficult to ascertain.

report results to stakeholders. The program of self- and donor-mandated evaluations could serve to elucidate what works and what does not and why; evaluations could serve as tools for oversight and accountability. The five-yearly reviews provide an opportunity to examine the outcomes and impact of the overall program and to adjust policies and strategy. Finally, the Fund uses standard training-evaluation tools to assess those training programs that do not fit into TA platforms.

75. **Implementation of key aspects of the M&E system is a recent development, and therefore the contributions of this system to the relevance, effectiveness and impact of CD will become clearer over time.** As discussed above, RBM and the CEF are only now starting to be widely used, and therefore, until recently, they did not contribute to prioritization, to the design of activities, or to monitoring and reporting, which has limited the effectiveness of the M&E system in enhancing learning and accountability. On the other hand, the system is already providing adequate information to facilitate reporting to donors, and it is expected that RBM and the CEF will be fully functional and able to provide information for the preparation of the 2023 review of CD.

76. **Overall, the design of the M&E system that has evolved over the past two decades is better aligned with the Fund’s requirements, but it needs to be further adjusted to respond to continued weaknesses and evolving CD strategic directions.** The focus and coverage of the M&E system was appropriately modified to focus on results, to cover all CD independently of how it is financed, and to treat country-tailored training as TA. But the system still needs to be adjusted to other strategic needs. First, the Fund needs an integrated evaluation strategy, which brings together the multiple evaluation tools and more strategically, strengthens quality control, enhances dissemination, and applies the results to drive change. Second, there are no standards or processes to assess CD results in the context of the Fund’s overall engagement with a member country, nor the extent of integration of CD with surveillance and lending, or the synergies between them. Adjusting the system to assess these country-level results will be a challenge, since currently the IMF does not have an evaluation system for surveillance and lending commensurate with that for CD. Third, a framework needs to be developed to assess CD activities focused on global public goods and other strategic priorities of the Fund, such as climate change, gender, and AML/CFT, that require a different evaluation perspective since they are not guided by a country’s demand and cannot be assessed by their contributions to a specific country. Fourth, ICD needs to find ways to strengthen staff commitment to RBM and to ensure that its results are used appropriately in the evaluation context—and address tensions with use of RBM for project management.

IV. EVIDENCE ON THE EFFECTIVENESS AND IMPACT OF IMF CD

77. **This section examines the effectiveness and impact of IMF CD activities.** IMF CD has been widely recognized for the high technical quality of its outputs—its TA reports are greatly appreciated by authorities as trusted advice, and its training is seen as best in class on macroeconomic and other policies close to its core mandate (see Enoch, 2022; and Pedraglio and Stedman, 2022). Over the past decade, the IMF emphasized achieving results on the ground in

addition to the quality of the CD outputs. It has made significant efforts to work with authorities on implementation of CD recommendations and focusing on effectiveness and impact. In the discussion here, effectiveness and impact are defined according to the CEF. Effectiveness relates to the extent to which a “CD project achieved, or is expected to achieve, the intended results envisaged in the RBM log frame;” while impact refers to “the CD project’s contribution to the Fund CD’s strategic goal of helping member countries achieve macro and financial stability and inclusive growth by strengthening the institutional capability of their agencies/governments to implement sound policies” (see 2020 Updated CEF, paragraph 6). The CEF also offers a more “IMF-centric” definition of impact that looks at “the CD project’s intended or unintended effects on the Fund’s strategic priorities (e.g., inequality, climate change, gender) in the country.”

78. **There is no single source of evidence that can provide a comprehensive assessment of the outcome and impact of IMF CD, but this section presents evidence from a variety of sources that together give a broad picture of its contributions and, to some extent, of results on the ground.** As explained in the previous section, the IMF is starting to monitor and assess the results of its CD. However, so far, none of the systems in place gives on its own the full picture of the IMF CD contributions, outcomes, and impact.⁴⁹ Still, this paper presents information from a variety of sources on different aspects of the results of IMF CD, that when looked at together provide a convincing picture of significant overall achievements. The following sub-sections examine respectively the data and ratings from the RBM system; the findings from project assessments, external and internal evaluations; the conclusions from country cases, the results of a stakeholder survey prepared for the IEO evaluation, and findings on training.

79. **The evidence presented below strongly suggests that Fund CD has furthered the objectives of most projects.** It is clear that many projects have contributed to improving the technical capabilities of the recipient agencies, helped strengthening the capacity of governments and agencies to implement sound policies, and made progress toward the outcomes and objectives. However, there is little information on the sustainability of these improvements nor on their impact, particularly at the country level.

⁴⁹ The Fund does not conduct impact evaluations of its CD activities in a systematic way, and the implementation of RBM and of more focused evaluations are too uneven to capture the full impact. Randomized controlled trials (RCTs) and other impact evaluations are expensive, difficult to structure, and unlikely to provide guidance while issues are still relevant. RCTs would require designating a non-intervention agency or country to serve as a control to be able to measure the differences in outcomes between with and without IMF CD. Even in the unlikely case that the Fund could identify two countries that are sufficiently similar, it would not be feasible for the Fund to purposely undertake not to provide similar CD to the control country during the assessment period. Also, within a country, it would be difficult to find agencies that produce similar outputs or services to serve as controls. Impact evaluations would take a long time to be implemented and by the time they are completed, their results would likely be overtaken by other changes.

A. What do RBM Data Tell About the Effectiveness of IMF CD?

80. **This section uses several indicators based on RBM data to measure the effectiveness of IMF CD.** The best indicator of a CD project's effectiveness in the RBM framework would be an average of the ratings of a project's objectives, weighted by their importance to meeting the project's goals. However, objectives are not usually ascribed weights and the IEO is not aware that such an average indicator is being calculated to get an overall picture of effectiveness. Moreover, even if weights were available, there were only about 40 projects completed between 2015 and 2020 for which all objectives had been rated, less than a quarter of the 174 completed projects. Instead, this paper uses four indicators based on non-weighted averages of objective and outcome ratings (each of these rating indicates the extent to which an aspect of the project met its goals). This is similar to the approach used by IMF staff that uses for its analysis of CD effectiveness the unweighted average of available outcome ratings for completed projects with at least one rated outcome—the indicator for which there are the most observations.

81. **Four indicators are used in the analysis below that give a plausible assessment of CD effectiveness:**

- First, the average of outcome ratings of completed projects (the indicator used by IMF staff). For the period between 2015 and 2020, there were 131 completed projects that had at least one rated outcome (about 75 percent of completed projects), and there were 935 rated outcomes in these projects.⁵⁰
- Second, a similar indicator but based on the average of objective ratings—there were 47 completed projects with at least one rated objective, accounting for 138 objectives in total.⁵¹
- Third, the average of a project's outcome ratings for all 131 projects with at least one rated outcome.⁵²
- Fourth, the average of a project's objective ratings (similar with the indicator mentioned above, but with uniform weights and for all 47 projects with at least 1 rated objective).

⁵⁰ CD projects included in the RBM database had on average 11.2 outcomes but completed projects, with at least 1 rated outcome, had an average of 7 rated outcomes. These figures suggest that outcomes were sometimes discarded during project implementation because they proved less relevant than expected when the project was launched or because they were difficult to assess.

⁵¹ CD projects included in the RBM database had on average than 5.6 objectives, but completed projects had an average of about 3 rated objectives.

⁵² This indicator differs from the first one in that an average rating is first calculated for each project by averaging all the project's outcomes. These averages are then averaged across all 131 projects. The two indicators would yield the same results if all projects had identical rating distributions.

82. There are potential shortcomings to each of the aggregate outcome and objective ratings, but there is tentative evidence that they are not seriously biased (Figure 2). A

concern about outcome based indicators is that they have a built-in shortcoming in that project managers can add, delete, or modify outcomes and outcome indicators during project implementation as they learn more about the recipients' implementation capacity and changes in circumstances.⁵³ Also, interviewees and some evaluations mentioned that sometimes highly rated outcomes do not translate into a successful project.⁵⁴ Unfortunately, there are still very few projects with rated objectives.⁵⁵ Also, such ratings could have an upward bias because project managers have discretion to adjust calculated objective ratings when they perceive a disconnect with the achievements on the ground. These potential shortcomings notwithstanding, objective based indicators are only slightly higher than those based on outcomes and there is a high correlation between outcome ratings and objective ratings, albeit in the small sample of 47 completed projects with rated objectives.⁵⁶

83. More than half of completed CD projects that have rated objectives fully or largely achieved these objectives. The figures and tables below look at the results of completed projects from different perspectives, and all of them suggest significant achievements of Fund CD.

- Figures 2 and 3 show the distribution of ratings for the different indicators for completed projects.⁵⁷ The figures show that the distribution of ratings is concentrated between partially and largely achieved and skewed towards better outcomes and objectives.
- Table 4 shows the averages of all ratings according to each of the four indicators for completed projects. These averages range from 2.50 for the outcome ratings of individual projects to 2.75 for the overall objective ratings. The average rating for all outcomes was 2.63, which is equivalent to 54 percent of the ratings distribution.
- An alternative perspective, in line with the practice in other IFIs, is to divide project results between satisfactory and unsatisfactory depending on whether their ratings (or rating

⁵³ In about 10 percent of projects outcomes are modified, and in another more than 10 percent of projects the end date is delayed (Board presentation by ICD staff—August 2021). Also, as mentioned above, it appears that many outcomes are deleted during implementation.

⁵⁴ For example, the 2017 evaluation of the MCM multi-topic technical assistance on central banking in Zambia, delivered in cooperation with Norway, indicated that "The outcomes of the TA were successful in several areas, but key objectives were not met." (MCM, 2017; paragraph 30, in Annex IV)

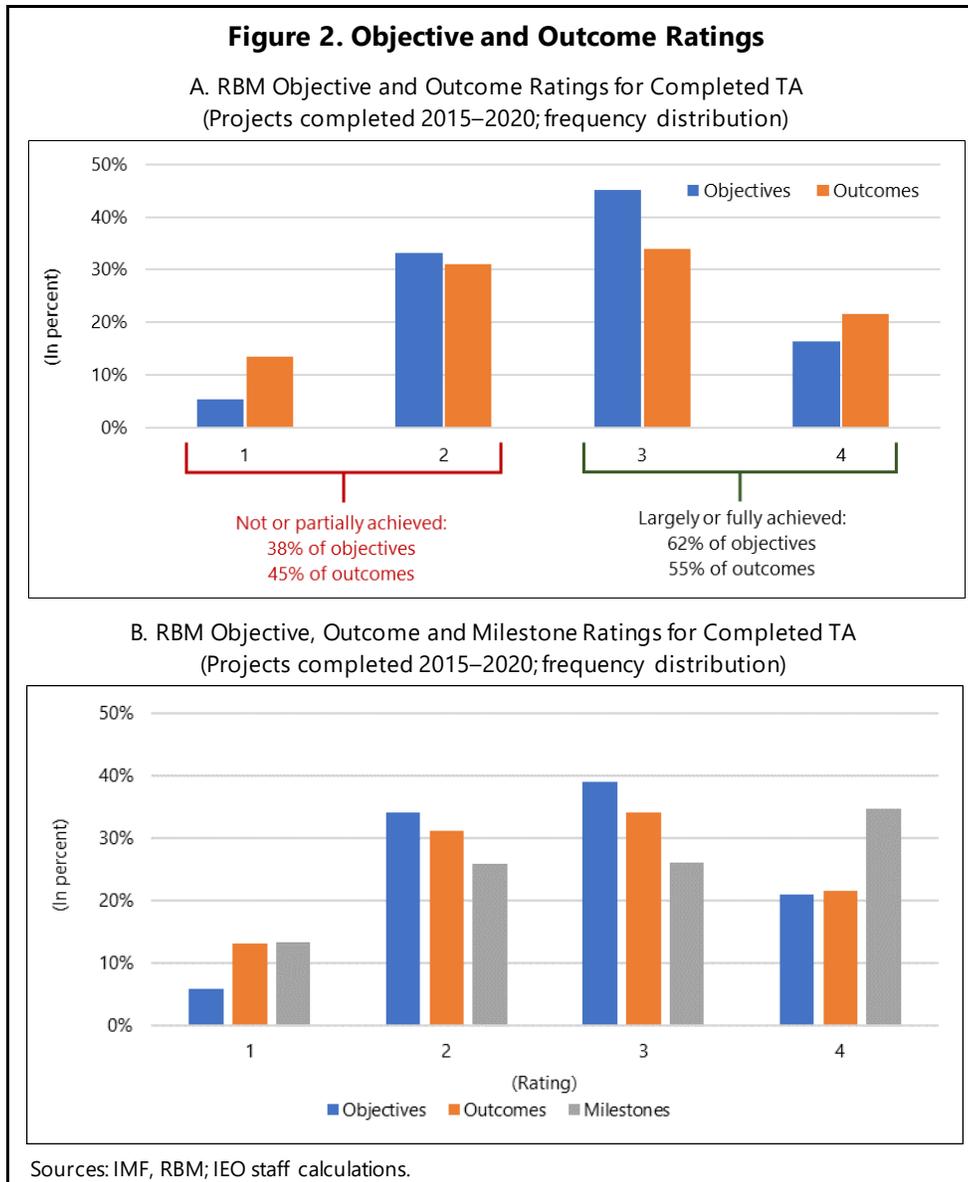
⁵⁵ In fact, all the performance indicators (i.e., both objective and outcome based, by project and overall) could also be calculated for each year and CDD. Some of these calculations are presented in Annex II, but the estimates are not reliable, given the small number of observations for each category.

⁵⁶ A regression analysis of objective ratings on outcome ratings found a significant coefficient of 0.7 and a positive intercept of 0.8, indicating that within the sample of 48 projects, managers pushed objective ratings upwards relative to outcome ratings, but not by a large amount.

⁵⁷ Figure 3 shows the distribution of the average rating per project, which have a continuous distribution that ranges from 1 to 4, and that were sorted into the corresponding three categories.

averages) are above or below the mid-point of the ratings range, in this case 2.5 which is the mid-point between ratings 1 and 4. Table 5 shows that the share of successful projects range from 52 percent when measured by objectives at the project level to 60 percent when measured by overall outcomes of all completed projects, with an average of 56 percent.⁵⁸

All in all, 55 percent seems a plausible and robust point estimate for the share of completed projects that fully or largely achieved their objectives.



⁵⁸ These ratings appear low when compared with the accomplishments in the completion reports, evaluations and country studies discussed below, but they are in line with the ratings presented in those reports.

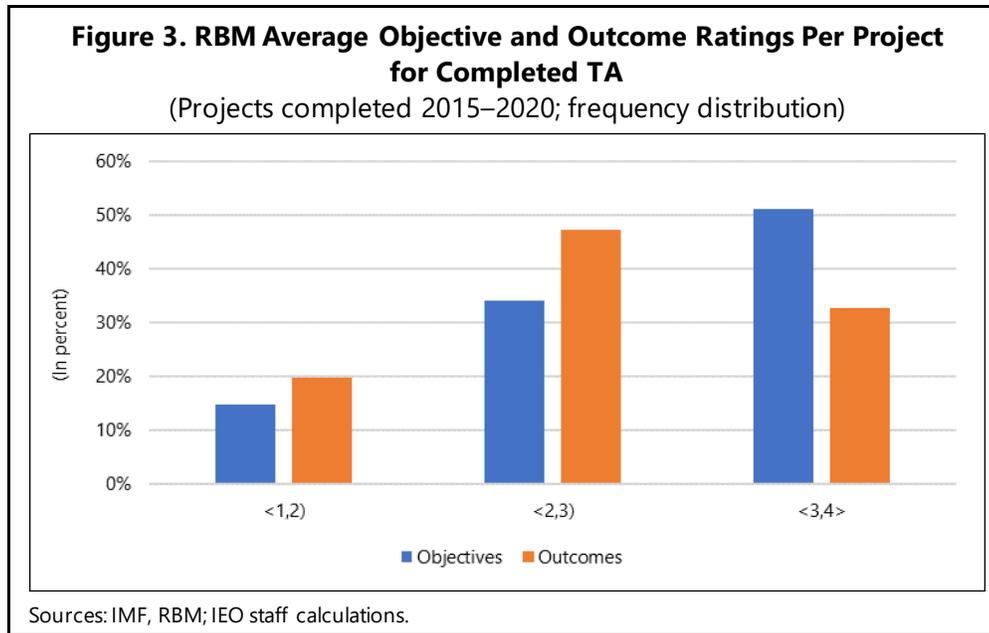


Table 4. Completed Projects—Average Ratings
(Projects completed 2015–2020)

Indicators	Average
All Objective Ratings	2.75
All Outcome Ratings	2.63
Objective of Individual Projects	2.65
Outcome of Individual Projects	2.50

Sources: IMF, RBM; IEO staff calculations.

Note: Each line is based on the corresponding definition described in paragraph 81 above. Based on a four-point scale, 1–4.

Table 5. Rate of Satisfactory Completed Projects
(Alternative indicators, projects completed 2015–2020)

Indicator	Satisfactory TA projects (In percent)
Rate of overall satisfactory objectives	60
Rate of overall satisfactory outcomes	55
Projects with satisfactory objectives	56
Projects with satisfactory outcomes	52

Sources: IMF, RBM; IEO staff calculations.

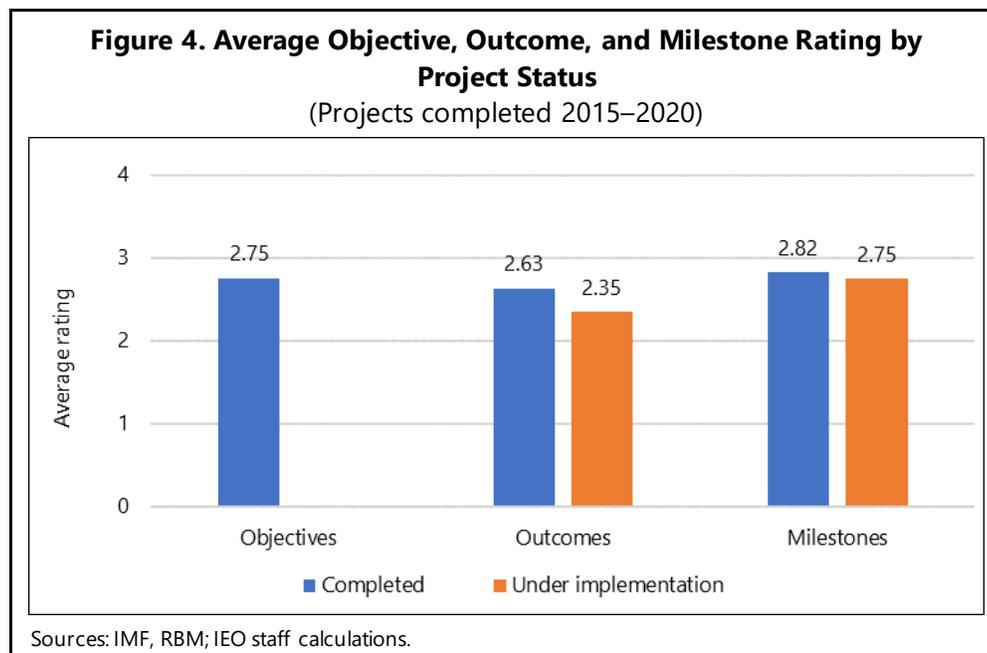
Note: Each line is based on the corresponding definition described in paragraph 81 above. Ratings above the mid-point of the rating distribution, i.e., larger than 2.5, are considered satisfactory.

84. **A similar picture arises when measuring effectiveness in terms of milestones, but milestone ratings have a weaker association with project performance than do objective and outcome ratings.** The database of milestone ratings is much larger than those of objectives and outcome ratings, which would be an advantage in using a milestones-based indicator.

However, milestones generally measure actions by the CD recipient and not necessarily outcomes. Moreover, project managers modify (and sometimes set) them during project implementation to use them as management and monitoring tools. Naturally, milestones' ratings rise over the life of projects; by project completion, milestones are generally rated as having been implemented. Modifying milestones makes them a more useful tool to monitor and incentivize implementation, but it reduces the evaluative value of their ratings at completion. In interviews, several experts indicated that it is common to have a disconnect between progress in meeting milestones and progress in outcome indicators and achievement of objectives. Many factors could lead to this disconnect including changing circumstances, a misspecification of the links between milestones and outcomes and objectives at project entry, or that recipients were more committed to the actions supported by CD than to its intended objectives. In any case, it is reassuring that milestone ratings show a similar picture as objective and outcome ratings. Table 6 and Figure 4 show that milestones average ratings are higher for completed projects than for those under implementation, and higher than objective and outcome ratings.

Project status	Average rating		
	Objective	Outcome	Milestone
Completed	2.75	2.63	2.82
Under implementation		2.35	2.75

Sources: IMF, RBM; IEO staff calculations.



85. **Over the next few years, as the RBM database grows, it will allow comparisons of TA effectiveness across units and themes within the Fund, how it evolves over time and how it relates to country characteristics and delivery modalities.** So far, there is not enough data to have a meaningful analysis of how the effectiveness of TA is evolving over time because the historical database is too small and incomplete. Until 2017, only externally financed projects were included in the RBM database and only projects still being under implementation were transferred to CDMAP. In a few years, however, it should be possible to track how the effectiveness of TA is evolving by examining time series of outcomes and objective ratings. Also, although there are significant differences in average ratings across CDDs, beneficiary countries, workstreams, and between projects managed by headquarters or regional center staff, there isn't yet enough data on ratings of completed projects for these comparisons to be robust (see Annex II, Table A.II.9).

86. **It is difficult to benchmark the RBM ratings of IMF TA against other IFIs because they do not produce similar indicators for TA projects, and because their TA activities are often very different from those of the Fund.** Most MDBs prepare project completion reports for their lending projects, which sometimes have TA components, and usually include outcome ratings at the project level. These outcome ratings reflect the judgement of individual evaluators and are informed, but not determined, by RBM data (in those organizations that have an RBM system). Every year, these individual outcome ratings are aggregated into an indicator of the share of projects whose outcome was satisfactory (the upper half of the ratings), which becomes the headline indicator of the institution's annual performance (see "Results and Performance of the World Bank Group 2021," IEG, 2021). This report also presents disaggregated ratings by regions, sectors, and WBG lending units, but it does not include a rating or discussion of outcomes of TA or training separate from the other objectives and outcomes of the project.

87. **The International Finance Corporation (IFC) (the private sector branch of the WBG), and the Asian Development Bank (ADB) rate different aspects of TA.** The 2021 Resource Allocation Plan (RAP), mentioned above, presents data on the development effectiveness of IFC advisory services, the closest indicator to an outcome of a TA project. The report mentions that after falling to an historical low of 38 percent projects with satisfactory ratings in FY2015–2017, ratings had been improving and reached 52 percent in FY2017–2019. In 2014, the ADB conducted a review of the role of TA in its operations. Based on a review of project assessments for 2007–2012 it found a success rate of 72 percent.⁵⁹ Caution is needed in comparing the percent satisfactory rates of IFC and ADB with the similar satisfactory outcome rating at the IMF. The IMF indicator is an average of RBM outcomes, while the IFC and ADB indicators are the average of project assessments. Also, the goals and characteristics of the TA projects are very

⁵⁹ The 2014 ADB evaluation (ADB, 2014) refers to an earlier "Special Evaluation Study on Performance of Technical Assistance" that found that the corporate information systems did not generate enough information to monitor and manage the TA portfolio properly, that staff and managers were held accountable for quality at entry and supervision but not for implementation and results, and that ADB was not building a database of lessons that would be useful to improving the use, effectiveness, efficiency, and results of its TA.

different. The ADB and other MDBs have put in place systems to calculate and publish RBM data on TA in future years.

B. What did Evaluations Find About the Effectiveness and Impact of CD?

88. **This subsection presents the main findings on effectiveness and impact from three evaluation products: project assessments, external evaluations, and internal self-evaluations by CDDs.** They provide an in-depth, but incomplete, picture of the effectiveness of particular aspects of IMF CD. It is an incomplete picture because for the most part the selection of what to evaluate was driven by the requirements of donors and prepared for projects financed by them, and not by an evaluation strategy aimed at assessing overall outcomes nor the main aspects of Fund performance.

Final Project Assessment Reports

89. **The IEO reviewed 10 final project assessment reports (FPARs) covering CD projects in countries with very different development levels and institutional capacity and led by all CDDs.** This represents more than half of the FPARs prepared since 2017 and therefore their assessments can be considered representative.⁶⁰ For each FPAR, the IEO review focused on the sections on “Results Overview and Key Accomplishments” and on the Logical Framework (including the description of results and the corresponding RBM ratings).

90. **Overall, the FPARs’ narrative points at significant accomplishments in building capacity and progress in achieving results on the ground.** Below are examples of achievements from two FPARs from each leading CDD:

- **Legal.** The 2018 FPAR for the project in Bolivia to strengthen the country’s AML/CFT framework developed a regulatory framework that is in line with the international standards that is being used for offsite and onsite supervision conducted using the risk-based tools, systems and procedures developed under the project. The overall assessment for this project was that milestones were achieved, Outcomes were largely achieved, and objectives were partially achieved. The 2020 FPAR for an AML/CFT project in Liberia indicates that the Liberian authorities enacted legislation and regulations on terrorist financing in line with Financial Action Task Force (FATF) standards, and that the Central Bank conducts offsite supervision of banks using tools developed under the project. Overall, the FPAR rated milestones and outcomes as partially achieved, as other goals of the project were not met.
- **MCM.** The 2020 FPAR for a project in Nigeria aimed at strengthening regulation and supervision indicates that with support from the project the Central Bank developed and implemented supervisory guidelines for estimating Pillar 2 capital requirements, risks,

⁶⁰ One caveat is that until 2021 FPARs were only prepared for projects supported by donors.

and stress testing. It also established guidelines to enhance the quality of risk analysis and internal controls in Nigerian banks. Most milestones were rated largely achieved and outcomes partially achieved. The FPAR also saw a high probability that the COVID-19 pandemic would put pressure on the performance of banks and on the feasibility of the Central Bank to implement its new systems and processes. The 2020 FPAR for a project in Belarus aimed at modernizing the monetary policy framework indicates that with project support the National Bank upgraded its forecasting models and that these had been utilized, and that it adopted new publication practices that led to greater transparency in policy making. Most milestones and outcomes were fully achieved, but it is noted that risks were high, from the pandemic and from political developments.

- **FAD.** The 2020 FPAR for a project in Guinea aimed at modernizing the tax system indicates that the authorities incorporated recommended changes in the Tax Code and implemented reforms in the personal and corporate income tax rates. All milestones and outcomes are rated largely or fully implemented. The 2020 FPAR for a project in Poland aimed at supporting a budgetary system reform indicated that with advice from the project, the Ministry of Finance had developed a Chart of Accounts in line with international standards. The FPAR noted that the project had moved slower than expected because of the COVID-19 pandemic, but that authorities were expected to implement the reforms and that a follow up project had been agreed. All milestones were rated largely or fully achieved, but there was no rating for the outcome or the objective of the project.
- **STA.** A 2019 FPAR describes the achievements of a project that supported improvements in data quality and dissemination in Albania, Bosnia and Herzegovina, Kosovo, and Montenegro. All the project goals were fully achieved, and the four countries implemented the Enhanced General Data Dissemination System. A 2019 FPAR describes the achievements of a Government Finance Statistics (GFS) aimed at helping improve fiscal data in nine Asian countries. This was a complex project that lasted four years and supported countries with very different initial conditions and intended results. In Indonesia and Thailand, the project helped expand GFS coverage to state-owned enterprises. In Malaysia, Mongolia, and Philippines the project coordinated FAD to expand data coverage to strengthen public finance management. In Cambodia, Lao PDR, Myanmar, and Vietnam, that required intensive CD, the project helped bring fiscal data systems to an operational level that can produce data useful for decision-making and analyses. Most milestones and outcomes for all countries were rated partially or largely achieved. Ratings were on average higher in the more advanced countries. Notwithstanding the significant achievements, several milestones were not implemented in the intensive CD countries.

91. **The achievements described above seem very significant, particularly taking into account that the projects had an average cost of less than US\$500,000 per country.** In many cases, particularly in low-capacity countries, some outcome ratings appear low relative to the description of progress in the body of the report, particularly considering the low cost and short duration of the projects. This disconnect between the ratings and the narrative may be due to ratings being driven by overly ambitious goals. A similar, related, disconnect exists between the average RBM rating and the findings on outcomes from evaluations and country case studies described below.

External Evaluations

92. **Most external evaluations provide a detailed description of the corresponding CD activities, outputs, and programs, and more recently of outcomes, sometimes over several years.** The selection of evaluation topics is driven by the requirements of donors and not by the goal of getting a comprehensive perspective on results. However, these evaluations cover a significant share of CD activities since donor funding accounts for over 60 percent of the CD budget. Some evaluations describe outcomes on the ground, even when their key goal, naturally, is to assess the contributions of the corresponding fund or regional center which is easier to measure in terms of activities and outputs.

93. **The author reviewed the last ten external evaluation issued between 2017 and 2021, representing more than a third of the evaluation reports issued since 2012.** They cover activities on a wide variety of topics, funded by many different donors, and taking place in countries from most regions. The format, approach and quality of these reports was discussed in Section III. The discussion below focuses on identifying examples of CD outcomes, and when available assessments of impact and sustainability. The examples presented below are just illustrative, as most reports describe the results on the ground for more than one project, some of which had many components. In comparing the findings and conclusions of different evaluations, it is important to keep in mind that the definitions of effectiveness and impact differ over time and across different evaluations. Many of the projects included in these evaluations were designed before the mainstreaming of RBM, and therefore lacked baseline information, making difficult to assess outcomes and impact.

94. **The CAPTAC-DR TA is focused on financial sector regulation and supervision, central banking operations, tax and customs administration, PFM, and macroeconomic statistics.** The 2018 evaluation of CAPTAC-DR gives a picture of significant achievements, e.g., in customs administration in Costa Rica, in financial supervision in Dominican Republic, and in central bank operations in Guatemala, although the assessment of effectiveness was complicated because many of the projects lacked baseline data. An evaluation survey of CAPTAC-DR member country authorities indicated that the skills, knowledge and organizational changes induced by the projects had been institutionalized and that they were likely to be sustained. The work of the center was highly appreciated by the recipient authorities.

95. **The Regional Technical Assistance Center in East Africa (AFRITAC East, AFE) provides CD to East African countries with the goal to develop and strengthen their capacity for effective macroeconomic management.** The 2019 evaluation of AFE assessed the effectiveness of AFE's CD as very good (rated 3 on a 4-point scale). In Tanzania, AFE had supported improvements in tax processing, including VAT refunds, collection, and accounting systems, it also had supported the work of the Bank of Tanzania and the National Bureau of Statistics in research and forecasting. Bank of Tanzania officials view the Fund's TA model of leveraging a network of central banks as better than academic training; they pointed at the support for the financial stability board is an example of assistance that could not be acquired elsewhere. The Ugandan authorities cited AFE's CD on tax investigations through visits to their Kenyan and South African counterparts as particularly effective. AFE helped the National Bank of Rwanda implement consolidated supervision, a critical intervention because foreign-owned banks dominate the banking sector.

96. **The Regional Technical Assistance Center in West Africa (AFRITAC West 2, AFW2) aims to support English and Portuguese speaking countries in West Africa strengthen human and institutional capacity to design and implement sound macroeconomic and financial policies to promote growth and reduce poverty.** The 2018 evaluation rated overall effectiveness, impact and sustainability as modest to good across all countries and activities (with the exception of PFM activities). In Nigeria, IMF CD help set up a TSA in Kaduna State which had already yielded increased revenues. TA supported the Bank of Ghana introduce new instruments and reprice existing ones leading to enhanced monetary transmission channels. In Ghana, Liberia, and Sierra Leone, the training of audit trainers on computer assisted audit techniques (CAATs) and data matching projects led to increased revenue collection. Authorities were generally satisfied with the quality of the TA and indicated that it had already led to tangible results that were likely to be sustained because the built capacity had been institutionalized. These results and assessments are particularly noteworthy given the fragile conditions of two recipient countries, political changes in four others, and an Ebola outbreak during the evaluation period.

97. **The Somalia Country Fund (SCF) aimed at supporting the authorities' efforts to build key economic institutions and develop capacity to implement sound macroeconomic policies.** It was implemented by all CDDs and MCD between February 2015 and June 2018. Most of the TA was hands-on, closely working with the staff of the relevant Somali departments, even if most of its activities were delivered from third countries. The 2019 Evaluation of the SCF found that during the initial stages Fund CD assisted the government on basic activities such as preparing a budget and organizing a functioning banking supervision framework). By the end of the project in some areas the Fund was supporting authorities in preparing monthly reports on budget execution, implementing excise duties and consumption taxes, and conducting off-site and on-site supervisory examinations. Given the challenging security circumstances in Somalia, the achievements of the SCF are remarkable, but obviously their impact and sustainability are difficult to assess. The Fund's CD work in Somalia is an example of high-risk CD that the international community is very interested that the Fund pursue (see Chopra, 2022).

98. **The Tax Administration Diagnostic Assessment Tool-Trust Fund (TADAT) aims to help countries strengthen their tax systems to better mobilize domestic revenue in a sustainable and economically sound way.** Two evaluations were conducted in accordance with the program requirements, a mid-term in 2018 and a second one at completion of Phase I in 2020. TADAT issued 78 reports in 65 countries (from low-income to upper-middle-income) and 9 at subnational levels. Overall, the evaluations found that “the TADAT product is robust and is “fit for purpose.” The TADAT assessment is conducted in close cooperation with officials from different agencies and there are many examples where the processes and methods were incorporated into the practices of those agencies. The evaluation found that in most countries the diagnostic had already led to improvements. Among the many examples presented in the completion report: Bolivia strengthened taxpayer compliance management, Cote d’Ivoire digitalized core processes, El Salvador implemented a post-TADAT assessment action plan, Georgia refined the compliance risk management framework, Jamaica improved on-time filing and payment, Kenya expanded the tax base by identifying unregistered taxpayers, and Tunisia implemented a risk-based audit system. The TADAT framework was “imitated” by countries and jurisdictions in all income levels that did not have formal assessments, such as: Afghanistan; Buenos Aires (Argentina); Canada; Cyprus; Hong Kong; Namibia; the Nigerian states of Ekiti, Ondo and Osun; Papua New Guinea; and Puerto Rico. The sustainability of TADAT’s achievements is reflected in that authorities continue to use the tools and processes after the TA is completed.

99. **These examples illustrate that IMF CD fostered significant results across many sectors and themes; they also point at how difficult is to measure effectiveness and impact in a consistent manner across different types of CD and across countries with different levels of development and institutional capacity.** Effectiveness and impact are assessed against goals set at the start of a project taking into account initial country and sector conditions. But goals in different projects and project components are set at levels that have different likelihood to materialize. Ex ante, these differences are captured by the risk ratings of the project, but ex post when risks materialized, they are reflected in lower outcome ratings.⁶¹ Moreover, there is a lot of subjectivity in setting milestones and outcome targets when designing the project. Somalia, the AFRITAC, and the Middle East Regional Technical Assistance Center (METAC) evaluations show achievements of CD activities that were major in the context of those fragile states—and probably should have been rated higher than in the evaluation reports. On the other hand, the same achievements would not be considered significant in other contexts. This issue is particularly important if the Fund and donors decide to assign greater weight to past results in the allocation of CD resources.

⁶¹ Rating of project risk has not been done in a consistent manner. It would take a very significant effort to try to normalize the project specific risk ratings across countries and sectors.

Internal Self-Evaluations by CDDs

100. **Most internal self-evaluations include some discussion of outcomes and impact, even if their main goal is to derive lessons to improve future CD.** The IEO reviewed eight internal self-evaluations from FAD, MCM, and STA, issued between 2012 and 2019, and representing more than a third of the evaluation reports issued during that period.⁶² Most of the evaluations used in one way or another the OECD-DAC framework, even those prepared before the launching of the CEF. On the other hand, the vast majority of projects covered by these evaluations had been designed before the widespread use of RBM log frames. Therefore, none of them was able to rely on project ratings on milestones and outcomes to assess effectiveness. While some of the evaluations tried to retrofit an RBM-like framework, it is expected that future evaluations that will benefit from RBM information will provide a more accurate picture of effectiveness but not so much of impact. Below we present illustrative examples of outcomes from evaluations from different CDDs.

101. **A 2019 MCM evaluation examined the outcomes, impact and sustainability of a program aimed to build regional capacity in public debt management in Africa.** The evaluation focused on the two largest recipients of support, Ghana and Cote d'Ivoire during 2015–2018. The TA under this project had close synergies with the Fund's supported programs and with IMF's surveillance. The evaluation found that in Ghana the project helped to significantly improve the Ministry of Finance's capacity for public debt management, it enhanced effectiveness of the primary auctions and helped develop a functional platform for secondary trading. The project had a transformative effect on the Ghanaian domestic debt market which was likely to be sustained. In Côte d'Ivoire the project was moderately effective, but attribution of results was difficult as the Fund had collaborated closely with other TA providers. It supported the regular production of a medium-term debt management strategy which was a significant achievement given the country's balance of payments crisis and debt-related risks. Improvements in the sovereign debt portfolio management, included the successful issuance of domestic bonds with longer maturities and of Eurobonds denominated with longer maturities. The project's achievements appeared to be well entrenched and likely to be sustainable.

102. **A 2012 STA evaluation examined results of the TA and training in the compilation and dissemination of macroeconomic statistics in Peru during 1993–2011.** During this period, Peru benefited from extensive TA, a 2003 data Report on the Observance of Standard and Codes (ROSCs), and training of over 100 officials. Fund TA and training had helped authorities overcome many capacity constraints and contributed to Peru subscribing to the Special Data Dissemination Standard (SDDS) in 1996. Examples of results included the implementation of changes in national accounts base years, compilation of a national consumer price index, implementation of updated methodologies for GFS and Monetary and Financial

⁶² In Annex III, the reviewed evaluations are marked with an asterisk. During this period there were no internal self-evaluations conducted by LEG (that was the leading department in several external evaluations).

Statistics (MFS); and stronger inter-institutional collaboration. These impacts demonstrated the long-term nature of the interventions and their sustainability. With regard to training, the mission found that the authorities made good use of the staff who participated in various courses on statistics organized by the IMF. Key staff in the Central Bank of Peru (BCRP) and the National Institute of Statistics and Information (INEI) that work directly on the compilation and dissemination of macroeconomic statistics have benefitted from this training. The main data-producing agencies (and domestic data users) indicated that TA had been effective across statistical areas, that it had contributed to the dissemination of more and better-quality statistics.

103. In 2019, FAD conducted an evaluation of the advice on TSA and Cash Management in six Sub-Saharan African countries during 2010–2018 that had received significant TA.

The evaluation found that TA had contributed significantly to capacity development in the six countries, even if to different degrees. TA reports were of high quality and authorities were satisfied with the support they had received. The evaluation assessed the implementation of recommendations, progress towards outcomes in several cash management categories (e.g., regulatory and institutional framework, cash forecasting, and analysis of cash flows), and the impact on higher level macroeconomic indicators (where attribution is much less clear). It found that where a TSA was established, government deposits in the Central Bank had increased significantly. The evaluation highlighted the implementation of TSA by the Federal Government of Nigeria (FGN) with IMF support which is considered one of the more successful PFM reforms in Africa. The TSA was launched in 2012 in tandem with a new Integrated Financial Management Information System to provide the supporting functionalities needed by participating agencies. In 2016, the finance minister said that “The TSA has provided complete and timely information on FGN cash resources, improved operational control on budget execution, enabled efficient cash management, reduced bank fees and transaction costs, facilitated efficient payment mechanisms, and has also reduced the FGN ways and means requirement to bridge the budget funding gap.” Following the success of this reform at the federal level, several state governments had asked the Fund for support.

104. The quality of the analysis and presentation of internal evaluations was higher than those of the external evaluations, but their selection was not geared to providing a comprehensive assessment of results and their coverage of effectiveness and impact was more limited.

The main goal of most internal evaluations is to draw lessons on how to improve IMF CD. To this end, an analysis of the quality and implementation of recommendations and of organizational issues at the Fund may be more important than assessing impact and therefore received more attention. External evaluations are often used to justify the continuation of a donor fund or initiative, which is often based on a description of outcomes and impact. Aligning more closely the selection, design, and implementation of the two types of evaluations would likely allow to have higher quality products, and probably a deeper description of outcomes since Fund staff are generally more knowledgeable of what has been achieved and what was the Fund CD contribution.

C. What did the IEO Country Cases Find About the Effectiveness and Impact of CD?

105. **This subsection provides a synthesis of IEO's country case studies in terms of effectiveness and impact of IMF CD, which have the advantage of a longer time perspective to assess CD results.** The IEO conducted 19 country case studies on high and medium intensity CD users, from all ADs, ranging from fragile states to high middle income emerging markets, and including both borrowers and surveillance only countries.⁶³ Each case study describes the CD provided by the IMF, and presents an overall assessment of its relevance, coherence, and effectiveness, impact and sustainability. This subsection is based on a review of all the country cases. It first discusses the key findings of a sample of the country cases that help illustrate how CD operates in different circumstances, and then distills lessons on the outcome of Fund CD overall.

106. **Sri Lanka was a high intensity CD user with moderate effectiveness; CD design was closely integrated with the Fund programs that were in place for much of the period.** In many areas, IMF CD was effective in building stronger macroeconomic institutions and policy making capacity. However, in several areas, results were mixed, as commitment to reforms, which was linked to program conditionality, dissipated with the termination of the program and personnel changes. Among the examples of effective TA are changes in tax policy that boosted VAT collection; development of a model-based forecasting and policy analysis system to inform policy formulation; and the launching of a national CPI. But other CD recommendations that were agreed under the program but that were not widely owned were ultimately not implemented or reversed, e.g., the new fiscal rules framework and a new central bank act. There were also reversals in tax policy and administration reforms. This experience in Sri Lanka shows the limitations of using program structural conditionality as a lever to implement CD recommendations and other reforms for which there is limited country ownership.

107. **Georgia was a high intensity user of CD that was highly effective and had a lasting impact on the economic management and performance of the country.** Georgia is a small upper middle-income country with a good reform record that has led to sustained macro stability and growth. CD was well integrated into the Fund supported programs that it had in place for most of the period. The authorities were eager to receive IMF CD which they considered superior to what they received from other TA providers, they engaged in the activities, owned the recommendations, and were committed to their implementation. Examples of impactful CD were strengthening the frameworks for monetary policy and operations; better financial regulations and supervision; the passage of a new insolvency law consistent with international good practice; and greater fiscal transparency and fiscal risk analysis. There were, obviously, areas where TA traction was mixed as domestic ownership was weaker and where more flexible advice, consistent with local conditions might have helped, e.g., in revenue administration. Still, Georgia is an example of a country where CD and program work complemented each other, but where there was strong ownership of CD recommendations that did not require conditionality to ensure implementation.

⁶³ Countries include Albania, Brazil, Cambodia, PR China, Congo DR, Georgia, Guatemala, Indonesia, Jamaica, Liberia, Moldova, Nigeria, Peru, Saudi Arabia, Senegal, Somalia, Sri Lanka, Uganda, and Ukraine.

108. **Cambodia provides a good example of high intensity CD yielding significant results, even after the conclusion of a program relationship.** IMF CD made significant contributions to Cambodia's policy reforms, to strengthened macroeconomic management, and to efforts to develop capacity. Authorities appreciated the IMF's technical expertise, implemented recommendations on a wide front, and continued to request and utilize large amounts of CD after they concluded their program relationship with the Fund. IMF CD supported significant successes across many fiscal areas leading to large revenue gains and improvements in customs administration, and to strengthening expenditure budgeting, reporting, management, accounting, and public procurement procedures. IMF training helped build macro-fiscal analytical capacity. IMF CD has supported improvements to the monetary policy framework, monetary and FX market operations, and to the macro prudential framework, regulations, and supervisory capacity. A 2019 Financial Sector Stability Review (FSSR) is seen by authorities as a roadmap for future TA and reforms. IMF CD has also helped Cambodia address data gaps and improve data quality.

109. **Albania was a high intensity user of CD, which contributed to the upgrading of Albania's policymaking frameworks and institutions over the period under review, though with mixed results in some areas of public finance and questions about long term impact.** Albania is a small, middle-income emerging market economy. During the evaluation period it was engaged in IMF supported programs for only a few years. CD was largely effective in supporting the Central Bank on monetary policy, banking regulation and supervision and establishing macroprudential policies. On the other hand, sustainable impact remained uncertain in many areas, particularly on tax policy and on PFM where CD was not able to overcome serious governance issues. These issues and the lack of reform in other areas raises questions on the sustainability of the policy frameworks and outcomes.

110. **Jamaica was a heavy user of IMF CD where effectiveness was closely linked to program implementation.** Jamaica is a middle-income country that until a decade ago had poor record in implementing Fund-supported programs and CD. Since then, the record of effectiveness and impact of CD improved as the design of CD was driven by a shared understanding by the authorities and Fund staff of the priority policy and institutional reforms needed to ensure a successful implementation of Fund-supported programs. TA contributed significantly to the successful implementation of fiscal rules, and tax policy reforms and strengthened revenue administration institutions and procedures were reflected in a 5 percentage points increase in the ratio of revenues to GDP. CD was especially effective in strengthening the governance and operational capacities of the Bank of Jamaica in the conduct of monetary and exchange rate policies, and in the development of the market for government securities. CD contributed to significant improvements in the production of macroeconomic statistics, despite limited absorptive capacity in the Statistical Institute, and Jamaica began implementing the Enhanced General Data Dissemination System (e-GDDS) in 2017. Less effective were efforts to restructure the government wage bill, reforms to adopt Basel II standards, strengthen supervision of non-bank financial intermediaries, and reforming AML legislation and strengthening its implementation, for Jamaica to exit the FATF's grey list. Authorities appreciate

the training provided by ICD, but its impact is limited by the substantial staff turnover. The sustainability of achievements will be tested once Jamaica moves to a surveillance status for a prolonged period.

111. **Congo DR was a high-intensity user of IMF CD but with large ups and downs and with very limited results.** DRC is a fragile state that did not have a program engagement for most of the evaluation period and had long disruptions in surveillance and CD due to political instability and security considerations. At times, however, it was one of the largest CD recipients, mostly externally funded, to support capacity building for PFM and revenue administration, particularly related to natural resources management. Two key projects were helping officials to use FAD's Fiscal Analysis of Resource Industries (FARI) tool to improve fiscal projections and assess the fiscal implications of tax policies in the mining sector, and to support the preparation of a mining code to strengthen tax policy and administration. The implementation of FARI was not successful because it elicited political push back, and also because the revenue agencies did not possess the skilled human resources to monitor the activities of mining companies. The mining code was approved in 2018 after several years of CD support in the face of proceeded weak political commitment. Its introduction has led to significant increases in mining revenue. Finally, at the authorities' request, CD was directed to the implementation of FSAP recommendations. Some progress was made toward strengthening the banking regulatory framework and implementing a risk-based supervision system, but overall impact was limited due to high staff turnover and limited skills.

112. **CD to Indonesia was generally seen as of high quality and well-tailored to country circumstances, but outcomes and impact were mixed at least in part because implementation of recommendations has been uneven.** Indonesia has been a moderate user of IMF CD and a surveillance only country during the evaluation period. CD was effective in supporting financial deepening, strengthening macro prudential supervision and crisis management, including through improvements of stress-testing capabilities; enhancing supervision of financial conglomerates; and regulation of AMF/CFT risks. TA also supported the improvement of statistics and the use big data to develop an enhanced residential property price index, expand the coverage of its producer price index to include service sector activities. Also Fund training of relevant officials has had a lasting impact. Because of a complicate legacy relationship, Fund staff has been receptive to Indonesia's requests, even on areas where past recommendations had not been implemented, e.g., the 2017 medium-term revenue strategy project that was explicitly requested by the authorities. However, lack of strict implementation of CD recommendations does not necessarily mean that the authorities did not benefit from the Fund's advice in designing their own plans. This is sometimes the case, particularly in countries with strong domestic capacity and sustained good economic performance such as Indonesia.

113. **Brazilian officials were appreciative of the relevance, timeliness, and technical quality of the limited IMF TA that they received and indicated that it had an impact on their policies.** Brazil is a large middle income emerging market economy with high capacity

across the public and private sectors. During the evaluation period, it suffered from significant shocks and big political changes, but it maintained a relatively stable policy framework and remained a surveillance only country. It requested and received limited amounts of TA, particularly in the fiscal area, which the authorities take seriously as an input to their policy decisions. Preliminary indications suggest that TA in revenue administration has been particularly effective, supporting strengthening the taxpayer registry, developing a comprehensive strategy to manage compliance risks, and improving taxpayer services. TA has also supported the creation of a registry of approved public investment projects; the preparation of a comprehensive guide to cost-benefit analysis for proposed investments and training of officials in its use; a new procurement law to strengthen transparency and value for money in investment bids; and developing a more detailed and quantitative analysis of fiscal risks from PPPs. On the other hand, there are areas where limited progress has been made so far such as reforming subnational fiscal rules and the whole framework of intergovernmental fiscal relations, but these are politically sensitive where the Fund's technical advice cannot be the only consideration.

114. **China has high human and institutional capacity but has continued to request IMF CD because it appreciates its technical quality and international perspective.** It is difficult to assess the outcome and impact of CD to China because it is often requested as one of several valued inputs for formulating their plans. Still there have been instances of CD recommendations that were implemented and other occasions when the authorities indicated that Fund CD had significant influence on their plans. Reportedly, IMF TA contributed to the reform of the VAT (replacing earlier service business taxes), to the consolidation of personal income tax rates, the centralization of social security arrangements and their more effective integration within the tax system, and to strengthening capacity to analyze the economic impact of tax reforms. CD also contributed to tax administration with advice on legal frameworks, compliance strategies, personal income-tax administration, strengthened taxpayer services, auditing and investigations, and performance measurement. Staff considers that IMF CD contributed to the establishment of a well-designed TSA. CD on financial reporting and balance sheet issues (including accrual accounting) may have fed into policy actions by the authorities to improve provincial level budget reporting and address debt issues. Finally, Fund CD contributed to operationalize macro-prudential policy, for example the subsequent adoption of counter-cyclical capital buffers and the need for clear criteria to both trigger and turn-off macro-prudential measures.

115. **The case studies show that Fund CD has been effective in supporting capacity building and policy reform across a wide range of issues and country circumstances.** The findings of the case studies reiterate those from evaluations and the RBM system. In most cases, country authorities and officials in implementing agencies appreciate the interactions with TA advisors for their high technical quality and relevant experience. They trust the IMF quality assurance and are therefore willing to implement TA recommendations. Implementation generally led to the intended outcomes, and when sustained to the broader desired impact. The country studies also illustrate how IMF training has built capacity across the world on many technical and policy issues.

116. **The expectations and assessment standards of the effectiveness of CD need to be further calibrated to the different country and agency circumstances.** The country studies show that a strength of IMF CD is that the intensity and design of activities are generally adapted to the different country and agency circumstances. Similarly, log frames (i.e., milestones, outcomes, and objectives) are calibrated to some extent to the capacity level of different countries and agencies, but they need to be further differentiated to be able to capture realistic levels of implementation and outcomes. To assess the effectiveness of CD it is important to calibrate the expectations and assessment standards more explicitly to the recipients' levels of human and institutional capacity. The evaluation standards, including objectives and outcome targets, cannot be the same in a fragile state as in a well-functioning low-income country (LIC), and certainly should be completely different for a high-capacity emerging market economy.

117. **The case studies in high-capacity countries show examples of high impact CD, even if the IMF recommendations are not implemented as drafted in the TA reports.** Brazil, China, Peru, Indonesia, and to a large extent, Saudi Arabia have domestic capacity to develop reform programs—or can without difficulty engage private advisors. The fact that these countries engage with the IMF expert teams is already an indication of influence and effectiveness. The country studies illustrated cases when authorities implemented actions as the TA experts had recommended (probably after incorporating authorities' suggestions). But there are also cases where the authorities used the TA report as an input in the design of a modified plan. Both these situations are examples of impactful TA, given the country circumstances.

118. **On the other extreme, most CD in fragile countries yielded limited outcomes in the short run.** The cases of Congo, Liberia, and Somalia suggest that countries that need capacity building the most, are likely to be those where it would be most difficult to measure the outcomes of CD in the short run. Some of the CD in these countries may yield high returns, but it is usually difficult to ascertain which activities and when. There are important differences across these countries. Liberia had significant capacity before the civil war and CD has helped reconstitute it, although sometimes with limited success. In Congo, often projects stalled because there was limited reform drive. In Somalia, because all the institutions of the state had to be recreated from afar, IMF staff modest but realistic objectives, many of which were met.

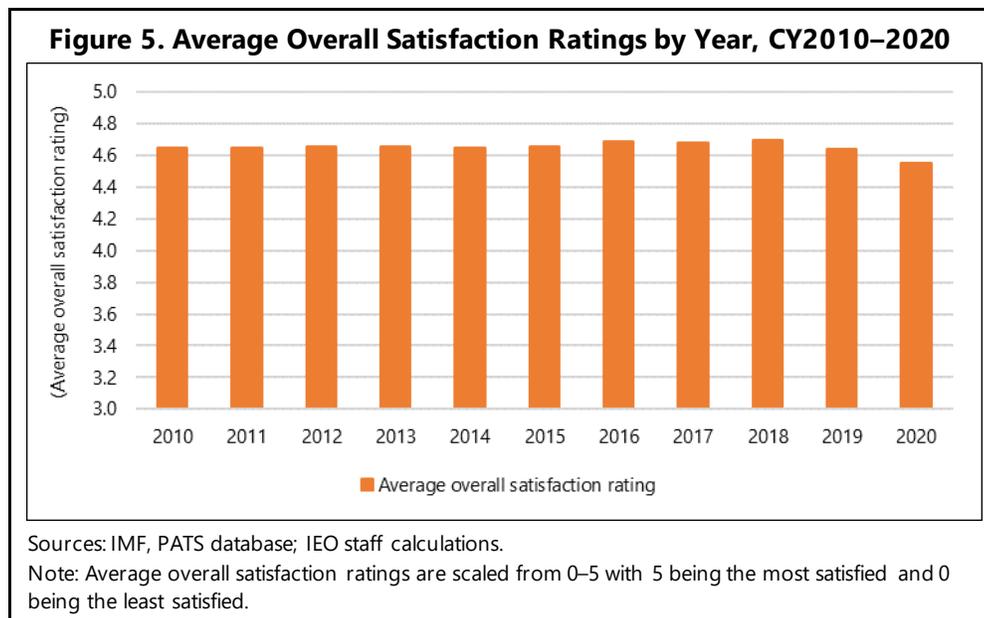
119. **The country studies show that the presence of an IMF supported program is not a silver bullet that would ensure the effectiveness of CD.** For example, comparing the experiences of Georgia and Sri Lanka, high intensity CD users with IMF-supported programs for most of the period, shows the limitations of relying on conditionality to foster ownership. Both countries made progress on CD recommendations during the program, when these recommendations served as the basis for structural benchmarks. However, their experiences diverged as soon as their programs were completed—Georgia, that owned the CD plans, continued to implement the agreed plans, while Sri Lanka backtracked on those issues on which it had no ownership. In Jamaica, another heavy user of CD where a Fund-supported program was in place for most of the period, the effectiveness of both CD and the program moved in tandem

reflecting changes in the authorities' ownership of both. The effectiveness of CD in Cambodia and Albania, that had programs for part of the period, continued even after the termination of the program relationship—with Cambodia making progress across all sectors, while the traction of CD was more narrow in Albania.

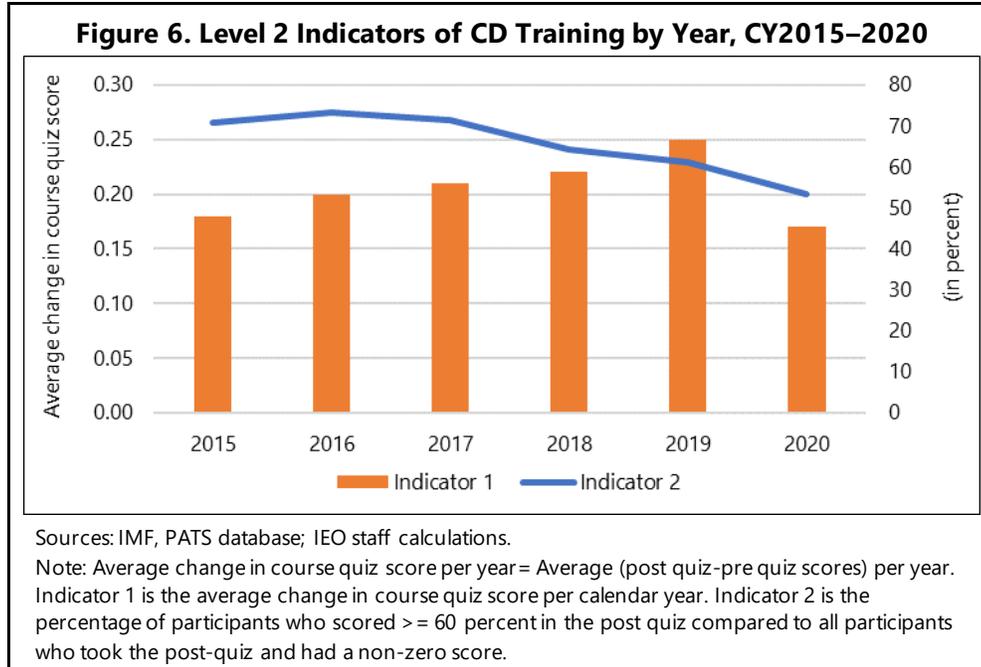
D. The Findings from Training Assessments

120. **End-of-course surveys give a very positive picture of participants' satisfaction.**

Level 1 surveys have had over the years high response rates and high ratings on different aspects of the course. The participants' average overall satisfaction rating was around 4.6 in a five-point scale throughout the past decade (Figure 5). Both the response rate and the level of satisfaction with the courses has fallen somewhat since the start of the pandemic as both the IMF and the participants were trying new engagement modalities.



121. **The comparison of pre- and post-course tests shows that participants learned the material, and many achieved a high level of proficiency.** During 2015–2020, more than 15,000 government officials that participated in over 600 courses took the Level 2 pre- and post-training exams (see PATS database). The post-course average test scores were 21 percentage points higher than the pre-course ones (from 49 percent to 70 percent) and over 70 percent of participants scored more than 60 in the final exams. These are significant learning gains that attest to the quality of the teaching and that participants were eager to learn the material—an indication of relevance. Ratings of the two indicators had been moving in opposite directions during 2015–2019, but both indicators fell during the first year of the pandemic as ICD and participants were trying new course arrangements (Figure 6).



122. **In surveys taken within two years after courses, participants and their supervisors gave very favorable feedback on IMF training.** In 2019 ICD surveyed participants in courses that had taken place between 2017 and 2018 and their supervisors to assess the effectiveness of its courses.⁶⁴ Almost 2000 participants and more than 600 supervisors responded to the survey, most of them from Central Banks and Ministries of Finance. Most participants, particularly those from countries in AFR and MCD, reported that the training had led to improvements in their analytical capabilities and job performance, and that it was useful for their current job. They indicated that course teachings were most commonly used to convey policy advice and to analyze economic and/or financial conditions in their countries or regions.⁶⁵ Supervisors' responses were similar to those of participants (e.g., that courses had impacted how participants did their job), but with differences in emphasis.⁶⁶ More than 60 percent of respondents worked in the same organization and same position following course participation—not surprising given

⁶⁴ See ROI Institute (2019).

⁶⁵ More than 80 percent of participants reported using the knowledge, concepts and skills presented in the courses in their day-to-day job functions, almost half of which reported using the knowledge and skills "a lot." This is reassuring because off-the-job continuing education often suffers of a high degree of "evaporation," i.e., participants forget what they learned within weeks of a course.

⁶⁶ The share of supervisors reporting that the participant was utilizing the course material is similar to that of the participants themselves—more than 80 percent, with almost half of supervisors indicating regular utilization. In particular, two-thirds of supervisors reported that participants used specific technical tools acquired during the course.

that less than two years had passed since the course.⁶⁷ Most participants reported sharing course knowledge with colleagues regularly (37 percent) or occasionally (49 percent), and this was also observed by their supervisors. Almost half of participants had contact with IMF staff post course—most of them in the context of Article 4 surveillance or of TA activities.

123. **Triennial surveys found that agencies in most member countries considered that IMF training had contributed to improve the effectiveness of their operations, a view that has not changed over time.** The 2018 triennial survey aimed to assess the effectiveness of ICD training programs and identify training needs. It was completed by 202 agencies from 114 countries.⁶⁸ Overall, more than 90 percent of agencies indicated that IMF training has helped participants do their jobs better and has improved policy formulation and implementation at the agency. Most respondents, across all income groups but particularly from LICs and EMEs, indicated that they used on a regular basis the skills and techniques learned at the IMF training. The two analytical skills most commonly cited by respondents from LICs and EMEs were financial programming and statistical compilation and dissemination. More than two-thirds of these agencies were also receiving TA, and most of them considered that training and TA were well-coordinated on country-specific issues. One-third of agencies believed that training had facilitated collaboration with IMF country teams on IMF policy issues.

124. **The very positive feedback received by all evaluation methods is a clear indication that Fund training has been effective in improving the work of participants and their agencies.** At the end of each course, participants were satisfied with how the training was delivered and they showed significant learning gains. This speaks of the relevance and technical quality of the training—some courses are considered the best and for some member countries the only ones available on their subject. An important contribution of IMF training that is seldom mentioned is that it has created a common language among central bank and government officials and other practitioners around the world—from national accounts to data standards, from financial programming to tax analysis. A shared language facilitates the dialogue and negotiations among policy makers. In fact, in the surveys, most agencies indicated that IMF training facilitated collaboration with the IMF. But beyond the relevance and technical quality, the surveys show that participants utilize the skills and knowledge acquired, and that their supervisors and agencies consider that these skills have improved the working of their agencies.

⁶⁷ Earlier follow-up surveys, including the one conducted in 2017, asked whether the IMF training had contributed to greater job responsibilities. Again, not surprisingly given the short, elapsed time, the vast majority responded in the negative. On the other hand, a large share of the senior-most officials in central banks in low- and middle-income countries have participated in IMF training.

⁶⁸ The 2018 triennial training survey was administered online from February 24–April 19, 2019, to 725 officials of government agencies that had sent three or more participants to ICD training courses from 2015 to 2017. It had a 29 percent response rate, as 202 agencies responded, more than half of them central banks or ministries of finance.

V. WHAT ARE THE MAIN DRIVERS OF EFFECTIVENESS?

125. **This section discusses the key drivers of CD effectiveness, i.e., the factors that affect the outcome of CD based on different sources.** The starting point is a modified categorization of the drivers of effectiveness proposed by ICD. These categories are explained and illustrated by providing examples from project assessments, internal, external and IEO evaluations, country studies, interviews with authorities and practitioners, and using RBM data.⁶⁹ Most of the discussion focuses on TA, but this is followed by a brief discussion of the drivers of effectiveness of training.

126. **At a Board presentation in 2021, ICD⁷⁰ highlighted three key drivers of effectiveness of CD over which there is general consensus: ownership, tailoring to context, and collaboration.** It found that CD projects are more effective when country ownership is strong, when CD is better tailored to country context, and when they exploit the synergies from internal and external collaboration. Similar findings had been identified by others, including the 2018 CD review and IEO (2005). While these categories are broad enough to accommodate many other characteristics that affect effectiveness, it may be helpful to consider a broader taxonomy:

- **Country or recipient characteristics**, including ownership, but also other country characteristics such as the level of development, fragility status, and absorptive capacity. An important characteristic is whether the country has a Fund-supported program or whether it is on a surveillance only mode. Over time, CD and other IMF activities can affect these country characteristics, but in the short term they are largely beyond the control of the CD team. Still, country characteristics are probably the main driver of effectiveness and should be considered in allocation decisions.
- **Design and delivery modes**, including tailoring, but also the selection of topics and themes, the complexity of a project, whether it is delivered by staff or LTX/STX, whether it is a single activity or part of a program, and whether it is managed from Washington or a regional center. These are CD attributes that have great influence on CD effectiveness and over which the project team has almost full control.
- **Interaction with beneficiaries, partners, and within the Fund**, including collaboration, but also other less intensive forms of interaction such as communication, collaboration, and cooperation. This category also covers how staff cooperates in integrating CD with surveillance and lending, and whether training is integrated with TA.

⁶⁹ Fund staff and academics have researched and published articles about how to assess and the drivers of results of TA; among these: Pastor (2009); Edison and others (2018); Bassanetti (2021); and Durand (2022).

⁷⁰ ICD Board presentation, August 31, 2021.

127. **These categories above provide a useful framework to think about effectiveness they influence each other.** For example, whether CD is integrated with lending and surveillance is linked to whether the country has a Fund-supported program. Some topics are more likely to elicit strong ownership than others, and the nature of the interaction with partners and beneficiaries depends on the level of development of the beneficiary and whether it is a fragile state.

Country or Recipient's Characteristics

128. **Ownership is universally seen as the key driver of effectiveness of CD and other development projects, as stated in the 2005 Paris Declaration on Aid Effectiveness.**

Shafik (IMF, 2018c) explained that “Whatever the nature of a country’s macroeconomic and development challenges, or its engagement with the IMF, it has come to be recognized that a country-led, demand-driven process with broad ownership is critical to ensuring success of IMF CD activities.” The effectiveness of CD projects and certainly the sustainability of its results depend critically on the recipient country and the corresponding agency regarding the project as “theirs” and being closely involved in planning, implementing and evaluating them (Keijzer and Klingebiel, 2019). The country studies found that CD was more effective and its outcomes more durable in countries that “owned” the CD agenda, e.g., Cambodia and Georgia. Moreover, the studies on Albania and Senegal show that within a country, projects that are most demand driven are more likely to succeed.

129. **Integration of CD with programs and conditionality increase effectiveness in the short run, but the evidence is mixed on whether outcomes can be sustained after the completion of the program.** There are many examples of synergies between CD and programs, with milestones being implemented as they had been absorbed as structural benchmarks into programs. For example, the self-evaluation on FAD Advice on Tax Policy in Africa found that implementation of recommendations was faster in countries where they were linked to a program; also in Jamaica, the TA recommendations that were implemented most consistently were those that were covered by the Fund supported program, e.g., tax reform and Central Bank policies (Ter-Minassian, 2021). In fact, Jamaica over the past few years is an example of a country where authorities owned and were committed to the implementation of both their program and CD, which generated positive synergies. Other times, conditionality was a substitute for ownership, still yielding results, although not always sustainable. Sri Lanka may be an example of a country where following the completion of the program the CD agenda stalled, suggesting lack of true ownership. The country studies in Senegal and Moldova found that officials were concerned that identifying CD recommendations with conditionality would undermine ownership and complicate the role of CD experts as trusted advisors. In interviews, LTXs indicated that some officials preferred water-down milestones to prevent them from becoming structural benchmarks. Based on regressions of RBM outcomes, Bassanetti (2021) and Durand (2022),

reproduced in Annex VI, found a correlation, though not robustly significant, between program and CD implementation.⁷¹

130. **There is no definitive answer to the question of how to build country ownership of CD projects and programs.** There is broad agreement that country-demand is a necessary condition for ownership, but it is less clear who can express this demand. In the current Fund set-up, ADs are charged with identifying country priorities and demands, which they do through consultations with central banks and ministries of finance, but not necessarily with line ministries and decentralized agencies. Closer integration with surveillance may help foster ownership of reforms, over the medium term, and therefore demand for the corresponding CD.⁷² The link between conditionality and ownership of CD is complex and needs further exploration, as it creates synergies as well as tensions. Building long-term relationships and trust and involving CD recipients in the design and evaluation of projects is helpful, but it does not guarantee that political and “turf” obstacles will be overcome. At the end of the day, ownership depends on social and political factors over which the IMF has little or no say.⁷³

131. **Countries and their agencies have a limited capacity to absorb CD, even when they are committed to the CD goals.** The recipient’s absorptive capacity, determined by its initial institutional and human development, is key to the ability of countries to implement CD and achieve sustainable impact. The 2018 Evaluation of CAPTAC-DR found that in many projects, including on National Income Accounts in El Salvador and PFM in Costa Rica, the institutional environment was the driving factor in the achievement of impact, more important than CD design and delivery. Similarly, the 2012 evaluation of STA TA in Peru found the support was effective because of the country’s strong capacity. The quality of the counterparts allowed the TA providers to pass on information and exchange ideas effectively and efficiently. It also noted that when domestic capacity was constrained, these tasks became more challenging. Sometimes absorption is limited by political push back to rapid reform. The Albania study (Everaert, 2022) found that “From the recipient side, better institutional capacity and staff retention promoted effectiveness, while political sensitivity of reforms proved an obstacle.”

132. **Fragile states need CD the most, but fragility weakens effectiveness and impact, at least if measured by the same yardstick as for other countries.** In general, low-capacity countries have a lower absorptive capacity, even if they can benefit most from implementing TA.

⁷¹ Bassanetti (2021) regressed RBM outcomes on a series of country and project variables. He found that “though not robustly significant, the correlation signs indicate that the CD recipient country’s ownership—proxied by the successful completion of the IMF program—or lack thereof—proxied by programs that went rapidly off track—can be relevant factors behind the probability of achieving the CD targeted outcomes.”

⁷² Surveillance could also benefit from integration, as CD could inform on the practical details of recommended reforms (see Ter-Minassian, 2022; and Legg and Sembene, 2022).

⁷³ The Georgia and other country studies show that staff is eager to support countries where it sees ownership and implementation capacity and where successful projects build their own momentum by generating greater willingness to supply additional CD (see Chopra, 2022).

The capacity of a country is usually correlated with its level of development. In a study based on RBM data and issued in August 2020, FAD found that Milestones' ratings are positively correlated with the recipients' level of the development and are lowest among fragile states. Similarly, Bassanetti (2021) found that fragile states have "a lower probability of achieving the targeted outcomes compared to other countries." More generally, beyond a certain level, additional CD may not lead better results. For example, there are indications that in Liberia pressures to deliver on the Capacity Building Facility led staff to significantly increase the intensity of CD delivery even though it was at odds with existing capacity constraints (Legg and Sembene, 2021). The studies on Congo, Liberia, and Somalia show that in fragile states it is particularly important to adjust the intensity, design and ambitiousness of projects to the limited absorptive capacity.

Design and Delivery Modes

133. **The effectiveness of CD depends to a great extent on its design and on how it is delivered.** Design starts with the selection of topic or theme, which includes whether the issue is within the Fund's core mandate and expertise, whether it refers to a global public good, and the inherent technical and political complexity of the issue.⁷⁴ A recurrent theme in evaluations and country studies is that authorities appreciate Fund CD particularly because it is focused on issues where the Fund is seen as having world class expertise. As the CD agenda expands beyond its recognized core areas, e.g., to macro-structural issues such as poverty or to global public goods such as climate change, the Fund will need to ensure that the CD advisors have high level expertise on the issues and that the new areas are recognized as well-grounded in the Fund's mandate. Another key design issue is that simple projects are generally more effective than complex ones—Bassanetti (2021) found that outcomes are negatively associated with the number of workstreams included in the project, suggesting that complexity weakens effectiveness.

134. **To enhance effectiveness, the Fund shifted towards programmatic CD delivered by experts located close to the recipients.** Over the past decade, the Fund shifted the majority of its CD from single mission activities to programmatic CD, and from Washington based experts to resident advisors and LTXs operating from regional centers (that now account for half of activities). RCDCs are highly valued by their beneficiaries as they facilitate faster responses and better tailoring programs to country circumstances. In fact, most country studies found that CD recipients prefer in country RAs who provide greater continuity and support in implementation, and the 2019 evaluation of the Africa Region Debt Management project found that continuity of advisors

⁷⁴ Even amongst core Fund CD topics, some are more politically difficult than others, e.g., reducing tax exemptions is likely to elicit more resistance than improving GDP statistics.

had been a key driver of effectiveness.⁷⁵ Bassanetti (2021) found that projects delivered through RTACs and by RAs are positively correlated with the probability of successful outcomes.⁷⁶

135. **The COVID-19 pandemic forced the Fund to shift to virtual delivery of TA and training.** In some cases, this shift was an abrupt acceleration of an ongoing process. There has been some slowdown in delivery, and a decline in RBM and training ratings. It is too early to know how much of the decline in rating is due to the delivery mode and how much to adjustment difficulties (and the global reduction in productivity). For activities, such as multi-country non-tailored training, a slight reduction in effectiveness may be more than compensated by a significant reduction in costs that may allow to expand deliveries.

136. **Would setting targets for the expected levels of successful outcomes and objectives improve the effectiveness of Fund CD?** It is an axiom in the management literature that setting targets for the achievement of corporate goals improves performance, i.e., it is a driver of effectiveness. Most corporations designate easy to measure key performance indicators, set targets for their annual achievements, and link compensation to achieving these targets. Promotions and compensation at the IMF are already linked to performance, but in a less formal manner and generally not linked to explicit corporate indicators. The Fund could set an explicit success target, e.g., 60 percent successful projects each year, as an aspirational goal. It is unlikely that such a goal by itself would lead to major performance improvements, but it may be helpful to focus staff on a specific goal. On the other hand, the Fund should be careful not to link meeting specific achievement levels to compensation—such a link is likely to lead to higher project ratings, but it risks creating incentives for to set less ambitious projects and for more lenient rating. Also, it would not be feasible to apply such a system only to CD but defining indicators of success in other Fund activities would be even more difficult.⁷⁷

Interactions with Recipients, Partners, and Within the Fund

137. **A key driver of effectiveness is close collaboration with beneficiaries in the design and implementation of projects including setting up milestones and rating performance.** An important step in that direction is to communicate clearly on goals and expectations, and to provide real time mutual feedback. The Fund can also help in facilitating communication between different participating agencies (STA, Peru, 2012). Fund staff can assist authorities set up mechanisms to coordinate donors' activities to ensure effective cooperation.

⁷⁵ There may be a risk that in some cases RAs and LTXs could become implementors rather than advisors.

⁷⁶ They indicated, however, that this positive correlation may be influenced by differences in the type or level of ambition of targeted outcomes between headquarters-managed CD and CD delivered through regional centers.

⁷⁷ The experience in other IFIs with the use of key performance indicators is mixed, as improvements in indicators do not always reflect similar trends in the intended goals.

138. **Cooperation between ADs and CDDs is critical for the successful integration of CD with surveillance and program work.** There have been cases of close and effective collaboration in the preparation of RSN and CSN that led to better cooperation and more effective delivery of CD.

139. **In regard to internal coordination, several senior staff involved in implementing CDMAP argued that a key success factor for CD was that projects should be managed by project managers.** These staff believe that TA projects will continue to grow in size and complexity and that they would require full time managers that consider project management their key accountability. Project managers would focus on financial management, reporting within the Fund and to donors, and relationship management. Many experts do not have the skills for these tasks, and they may prefer to devote their time and energy to the substantive component of the TA project. But creating a career stream of CD project management is not without risks. Recipients have indicated that a key element that they appreciate and prefer the IMF staffing model of CD activities, where staff are knowledgeable and have experience on the substance of the CD.

Enhancing the Effectiveness of Training

140. **The best way to improve learning and utilization would be for training to be delivered in-country to intact teams and followed up by hands-on TA.** In surveys and interviews, training participants and their sponsoring agencies suggest organizing training for teams of officials in their own institutional context. Courses should use more country- and region-specific materials and case studies. A step in these directions is to increase the number of participants from same country to create peer groups, and to provide in-country follow-up and support. The Funds efforts to integrate training with TA and to tailor courses to specific needs are steps in the suggested directions.

141. **Progressively, non-country specific, generic training can be delivered via distance-learning methods.** The pandemic has accelerated the Fund's efforts to expand the reach of its training by using teleconferencing and online courses. These methods allow officials to time and pace learning at their own convenience. This is an ongoing global trend and initial data suggests that any loss in initial effectiveness is more than compensated by cost reductions that permit expanding offerings. Also, distance learning facilitates the creation of virtual communities of practice that allows officials to learn from peers in a continuous manner. The Fund is already working on massive online open courses (MOOCs) and podcasts that serve to train officials and to educate people globally.

VI. CONCLUSIONS AND LESSONS

142. **This section presents this paper's main conclusions on the effectiveness of Fund CD and lessons to further enhance the M&E system.** The paper found that Fund CD has been effective in supporting capacity building and policy reform across a wide range of issues and

country circumstances. It also found that after a major overhaul during the evaluation period, the design of the M&E system is broadly adequate for the Fund's requirements, but it needs to be adjusted further to reflect evolving CD strategic directions and implementation needs strengthening. This overall assessment is based on the triangulation of evidence from RBM, evaluations, and case studies prepared by the IEO.

143. **The paper found that most CD projects met their goals at least to a large extent.** RBM data shows that more than 55 percent of the CD projects were rated as having had fully or largely satisfactory outcomes and objectives. In fact, the narrative of evaluations and country studies suggest a higher level of achievements.⁷⁸ Country authorities and officials in implementing agencies appreciate the interactions with TA advisors for their high technical quality and relevant experience, particularly on areas close to the Fund's core mandate and expertise. Fund TA and training is sought in high-capacity countries—an indication of its reputation and quality. The demand for Fund advice by these countries reassures those with weaker capacity to seek Fund advice and follow its recommendations. Many CD programs in fragile states showed major achievements considering their context, although impact and sustainability were difficult to assess.

144. **The paper identified drivers of effectiveness that could be helpful in the prioritization and design of the Fund's CD program.** As much as possible, the Fund should tailor simple projects to the specific country needs. Preferably the Fund should focus on areas of core expertise and coordinate with other CD providers to assist in other areas. If feasible, CD should be provided from a regional center or by a resident advisor; and in any case, the Fund should ensure continuity of the relationship and support in implementation. The intensity and ambitiousness of projects should be adapted to country circumstances, particularly for low-capacity countries and fragile states. CD should complement the reforms required by an IMF-supported program, while keeping a clear differentiation between conditionality and CD recommendations. The volume, scope and intensity of CD should be determined by the extent of a country's ownership of the reform agenda. The Fund should establish good communication with authorities to ensure that there is sufficient demand for projects before they are launched, and CD experts should work closely with the authorities in the design, implementation, and evaluation of projects, including RBM log frames, which could serve to foster and strengthen country ownership.

145. **Overall, the design of the M&E system is broadly adequate for the Fund's requirements, but it needs to be adjusted to reflect evolving CD strategic directions.** Over the evaluation period, the Fund took a number of major steps to strengthen its approach, and it now has a wide-ranging multipronged M&E framework that comprises four main elements: RBM,

⁷⁸ In responding to IEO's survey, more than 90 percent of recipients agreed that IMF CD provides high quality advice, that projects typically achieve their stated near-term objectives, that CD was effective in building my institution's capacity and that it had a sustained impact (see Question 11).

ex post evaluations, the five-yearly review of CD, and the evaluation of training. But the M&E framework needs to be adjusted further to take into account evolving strategic directions:

- Although the Fund is committed to demand-driven, country-centric CD, there are no standards or processes to assess CD results at the country level, neither on their own nor in the context of the Fund's overall engagement with a member country.
- The Fund aims at integrating CD with surveillance and lending, but there is no framework to assess the synergies between them. Addressing this gap is a difficult challenge because the Fund does not evaluate and surveillance or lending operations.
- There is increasing demand for the Fund to deliver CD activities focused on global public goods and other Fund priorities that require a different evaluation perspective.
- To address these challenges and guide its evaluation work, the Fund needs a strategic multi-year integrated evaluation plan, prepared and agreed with donors and RCDCs, and covering internal and external, country and thematic evaluations, and evaluations on all aspects of CD including delivery modes, funding, and governance.

146. **There are some implementation challenges that are common to all aspects of the M&E system, among them:**

- There is very limited **authorities' involvement** in monitoring and evaluation, even though this involvement could be key to develop ownership.
- There is no systematic **quality assurance** of evaluation reports and processes.
- There is little **dissemination** of evaluation lessons.
- Evaluation findings and lessons are not yet used in any systematic way in the **prioritization and design of CD activities**.
- There is little **monitoring of results following the conclusion of the Fund involvement** in CD projects.

147. **RBM provides a structure to help design results-oriented projects and to monitor progress during implementation, and to report results to stakeholders.** Looking forward, it is important to foster greater ownership of RBM by LTXs and other TA experts to prevent continued compliance problems or that RBM becomes a box-ticking exercise. Attention needs to be paid to avoid creating incentives for rating inflation since staff can justifiably believe that donor money will follow high rated results. There are no systems to distill and disseminate lessons from the growing RBM database.

148. **The quality of ex post evaluations has been improving since the introduction of the CEF, but there is still room to improve how evaluations are prepared and used.**

- **Project completion assessments** are high-quality streamlined documents, but they are not widely disseminated even inside the Fund. They could become the basis for a web-based knowledge management database of lessons and building blocks for broader evaluations.
- The quality of **external and internal evaluations** has improved with the introduction of the CEF. It is now unclear that there is a need to differentiate between them, beyond donors' fiduciary requirements, as both internal and external resources are deployed in support of most activities. Both products would benefit by having common planning, implementation, quality assurance and dissemination processes.

149. **The five-yearly strategy reviews of CD provided the opportunity to examine the overall CD program with some time perspective and facilitated incorporating lessons from experience into the process of adjusting policies and strategy.** These reviews focus on IMF policies and practices, but only marginally try to assess the outcomes and impact of CD. From learning and accountability perspectives, policy reviews would benefit if they were informed by a comprehensive assessment of CD effectiveness and impact.

150. **For evaluating most training courses, the Fund uses an appropriately modified version of the four levels assessment method commonly used by trainers.** It assesses participant satisfaction and learning. It also conducts interviews and surveys of participants and their supervisors to assess whether they utilize the skills and knowledge acquired, and whether these skills had improved the working of their agencies. The results of tailored training that is integrated into a country or regional TA program is evaluated as part of such program. ICD will need to develop a framework to evaluate distance learning courses.

ANNEX I. EXAMPLES OF VARIABLES IN A LOGFRAME TEMPLATE

(Extracted from CDPORT)

CD Dept	Workstream	Topic	Objective	Objective Description	Outcome	Outcome Description	Outcome Indicator	Outcome Indicator Description
FAD	Public Financial Management	Public Financial Management Laws and Institutions	Stronger PFM laws and institutions	Improved laws and effective PFM institutions	A more comprehensive legal framework covering all stages of the PFM cycle	A more comprehensive legal framework covering all stages of the public financial management cycle is enacted	Comprehensive PFM legislation (FTC 2.2.1)	Comprehensive PFM legislation (FTC 2.2.1)
MCM	Monetary, Exchange Rate, and Capital Account Policies	Model-Based Monetary Policy Analysis and Forecasting (MPAF)	Incorporating FPAS into the decision-making process	Developing a fully operational Forecasting and Policy Analysis System (FPAS) and incorporating it into the decision-making process	Monetary policy communication strategy strengthened	Transparency of decision-making process enhanced and monetary policy communication strengthened, increasing credibility of the CB	Monetary Policy (Inflation) Report is published at least 4 times a year	Monetary Policy (Inflation) Report is published at least 4 times a year
LEG	AML/CTF	Measures Against Tax Evasion: Legal Framework	The country strengthens its tax and related AML legal framework	The country strengthens its tax and related AML legal framework, thereby providing the legal basis to address risks of laundering of proceeds of tax crimes	The country's legal framework contains provisions on criminalizing tax crimes	The country's legal framework contains provisions on criminalization of tax crimes and AML measures to tackle laundering of proceeds of tax crimes	Laws on tax and AML measures in line with international standards are in place	Laws on tax and AML measures in line with international standards are in place to address proceeds of tax crimes
STA	Government Finance	Government Finance	Strengthen compilation and dissemination of fiscal statistics	Strengthen compilation and dissemination of fiscal statistics	Source data are adequate for the compilation of these macroeconomic statistics	Source data are adequate for the compilation of these macroeconomic statistics	Source data for annual GFS are available on a timely basis	Source data for annual GFS are available on a timely basis
ICD	General Macroeconomic Analysis	Financial Programming and Policies (FPP)	Strengthen analytical skills of a cohort of country officials	Strengthen analytical skills of a cohort of country officials, as a basis for better policymaking	Improved skills for macroeconomic policy analysis and forecasting	Improved skills for macroeconomic policy analysis and forecasting	A medium-term training strategy is developed and agreed with the authorities	A medium-term training strategy is developed and agreed with the authorities

ANNEX II. RBM INFORMATION ON IMPLEMENTATION, COMPLIANCE, AND EFFECTIVENESS

The database. This annex contains information about RBM coverage, compliance, and ratings of milestones, outcomes, and objectives for CD projects. It discusses compliance with defining and rating objectives, outcomes, and milestones. It is based on the RBM database as of end-December 2020, and includes projects started between 2013 and 2020 cleaned using the methodology in used Bassanetti (2021).¹ The database contains TA projects and tailored training activities and has over 6500 records with 34 variables each (on average more than 10 records per project, representing each individual objective, outcome, and milestone).

As shown in Table AII.1, as of end-2020, there were 577 TA projects covered in the RBM database, of which 176 were considered completed as of end-2020.² In addition, 400 projects were under implementation. Many older projects (both, completed and under implementation) were added to the database in 2017 and most of them were labeled as having started that year—which explains the big jump in projects initiated in 2017.

Start year	2013	2014	2015	2016	2017	2018	2019	2020	Total
Cancelled	0	0	0	2	3	3	2	0	10
Completed	11	26	10	25	68	29	6	0	175
Under implementation	0	3	15	26	174	85	62	27	392
Total	11	29	25	53	248	123	70	27	577

Source: IMF, CDPORT database.

Information in RBM templates. Projects were designed using a log frame that spelled out the project's objectives, the expected outcomes and implementation milestones. All the projects in the database had defined objectives and outcomes, and more than 95 percent had defined milestones (see Table AII.2). Of the 28 projects that did not specify milestones, only 6 had been started since 2018—three of which had been launched in 2020 and implementation was only getting started. The other projects missing milestones were started before log frames were mandatory and were retrofitted into the database in 2017. Compliance in using log frames to

¹ RBM data was collected in CDPORT format through FY22 when data collection was moved to CDMAP, but only information on projects with an end date after May 1, 2021 were migrated to CDMAP. ICD is working on creating a comprehensive database including older CDPORT information, but this will not be finished in time for this report. ICD staff cleaned the data base using the methodology used in Bassanetti (2021). The main corrections to the database were to delete information on multi-country training and non-tailored single country training, and categories of projects that were not yet under implementation (i.e., on hold and requested). Also, the analysis does not cover 10 projects started between 2016 and 2019 had been canceled by end-2020.

² In fact, only 116 projects were flagged in the system as completed. However, following Bassanetti (2021), 60 projects that were scheduled to be completed more than three months before the closing of the database were relabeled as completed under the assumption that most of them had been completed but that the project manager had not updated the "Completed" flag.

design projects was already high in 2020 and it is expected to be universal with full implementation of CDMAP.

Start year		2013	2014	2015	2016	2017	2018	2019	2020	Total
Completed	Count	9	21	8	17	66	29	4	0	154
	Percent	82	81	80	68	97	97	67		88
Under implementation	Count	0	3	15	26	174	90	62	24	394
	Percent	-	100	100	100	98	100	100	89	99
Total	Count	9	24	23	44	243	122	68	24	557
	Percent	82	83	92	84	98	99	97	89	95

Sources: IMF, RBM; IEO staff calculations.
Note: The complement are those projects without any defined milestone.

Milestones Rating Compliance. Project managers are expected to rate at least once a year the implementation of milestones and the progress towards achieving the expected outcomes. Rating of milestones is a key monitoring exercise and can serve to engage authorities in a dialogue on how the implementation of agreed actions is progressing. Outcome ratings are based on agreed indicators—estimating these indicators periodically and discussing them with authorities would help improve the results of the project or at least help understand what is not working well.

During 2016–2020 (the last five years in the database), the Fund launched 512 projects with an average of more than 27 milestones per project.³ Overall, about 78 percent of the about 14000 milestones were rated—a high proportion for a system that is still being phased-in, and this proportion is even higher for projects initiated in 2018–2019.⁴ Still, the data shows some significant challenges, e.g., only in half of the completed projects were all milestones rated and in 10 percent of the projects no milestone had been rated by end-2020.

Outcome Ratings Compliance. Table All.3 shows that almost three-quarters of completed projects have all their outcomes rated (and a slightly higher share had at least one rated outcome). Again, these figures suggest that RBM processes are taking hold. Eventually, with the full implementation of RBM within CDMAP it can be expected that rating compliance will be much closer to 100 percent.

³ The discussion on defining the logframe variables and on rating milestones is based on projects' starting year, as are tables All.1 and All.2. The discussion below on rating outcomes and objectives, as well as on effectiveness is based on the projects' completion year.

⁴ 2018–2019 were the first two years when all projects had to be designed within the RBM framework. During these two years the Fund launched 188 projects, of which 36 had been completed by 2020.

<i>Completed projects with at least one:</i>								
<i>Completion year</i>		2015	2016	2017	2018	2019	2020	Total
Objective rated:	Count	1	3	5	8	15	15	47
	Percent	50	33	36	32	27	22	27
Outcome rated:	Count	1	3	7	20	46	53	130
	Percent	50	33	50	80	84	77	75
Milestone rated:	Count	1	2	4	22	52	63	144
	Percent	50	22	29	88	95	91	83
<i>Completed projects with all:</i>								
<i>Completion year</i>		2015	2016	2017	2018	2019	2020	Total
Objectives rated:	Count	1	3	5	8	12	11	40
	Percent	50	33	36	32	22	16	23
Outcomes rated:	Count	1	3	6	20	44	50	124
	Percent	50	33	43	80	80	72	71
Milestones rated:	Count	1	1	4	21	47	49	123
	Percent	50	11	29	84	85	71	71

Source: IMF, CDPORT database.

Objective Ratings Compliance. Table All.3 also shows that only 27 percent of completed projects had at least one of their objectives rated (and only 23 percent had all of their objective rated). These numbers appear surprising low, particularly since objective ratings are, prima facie, calculated as a weighted average of outcome ratings. Another surprising finding is that compliance appears to have been (marginally) lower for projects started after 2017.

Effectiveness

Rating objectives and outcomes. Objectives and outcomes are rated on a 4-point scale, where 1 stands for "Not achieved," 2 for "Partially achieved," 3 for "Largely achieved," and 4 for "Fully achieved." Table All.4 presents the distribution of ratings for objectives and outcomes for completed projects for which there was at least one rated objective or outcome correspondingly. Data is disaggregated by completion year and totals are additionally expressed in percentages.

Table All.4. Distribution of Objective and Outcome Ratings						
<i>Objective ratings:</i>						
Columns: Objective rating	1	2	3	4	Total	
Rows: Completion year						
2015	0	0	0	1	1	
2016	0	1	1	1	3	
2017	0	3	5	2	10	
2018	1	2	6	6	15	
2019	2	28	21	4	55	
2020	5	13	21	15	54	
Total	<i>Count</i>	8	47	54	29	138
	<i>Percent</i>	6	34	39	21	100
<i>Outcome ratings:</i>						
Columns: Objective rating	1	2	3	4	Total	
Rows: Completion year						
2015	0	0	0	2	2	
2016	1	1	0	1	3	
2017	0	5	10	9	24	
2018	11	22	27	17	77	
2019	51	184	180	112	527	
2020	59	79	100	60	298	
Total	<i>Count</i>	122	291	317	201	931
	<i>Percent</i>	13	31	34	22	100

Source: IMF, CDPORT database.

Aggregated ratings of objectives and outcomes. Table All.5 presents the shares of Satisfactory (SAT) and Unsatisfactory (UnSAT) objective and outcome ratings, based on Table All.4. Objectives and outcomes are considered SAT when they are rated as fully or largely achieved (ratings 3 and 4 in the four-points scale).

	Objectives		Outcomes	
	Count	Percent	Count	Percent
SAT	83	60	518	55
UnSAT	55	40	417	45

Sources: IMF, RBM; IEO staff calculations.

Ratings of projects. Table All.6 presents the distribution of Satisfactory or Unsatisfactory projects according to both, their objective and outcome ratings. The data is calculated by first averaging the ratings of all the objectives (or outcomes) linked to each project. It then considered a project as having a SAT performance if it had an average objective (outcome) rating higher than 2.5 (the middle of the rating distribution). Table All.6 shows that a majority of projects had average ratings of objectives and outcomes that were rated largely or fully achieved (56.3 percent and 52.3 percent, respectively).

	Count		Percentage	
	Sat	Unsat	Sat	Unsat
Objectives	27	21	56.3	43.8
Outcomes	69	63	52.3	47.7

Sources: IMF, RBM; IEO staff calculations.

Average Ratings. Table All.7 presents the average ratings for objectives, outcomes, and milestones for completed projects and for those under implementation.

Project status	Average rating		
	Objective	Outcome	Milestone
Completed	2.75	2.63	2.82
Under implementation		2.35	2.75

Sources: IMF, RBM; IEO staff calculations.
 NB: The database includes ratings some project under implementation, and these average 2.66. This figure is not included in the table because RBM guidelines only require objectives to be rated at completion.

Alternative indicators of performance. Table All.8 presents indicators of performance across all completed projects. The first two lines are the averages for all objectives and outcomes, respectively (from Table All.7). The last two lines are the average of the average ratings for individual projects. Line three is based on objectives and line four is based on outcomes.

Indicators	Average
All Objective Ratings	2.75
All Outcome Ratings	2.63
Objective of Individual Projects	2.64
Outcome of Individual Projects	2.50

Sources: IMF, RBM; IEO staff calculations.

ANNEX III. COMPLETED PROJECT ASSESSMENTS BY YEAR AND CDD, 2012–2021

Number	Title	Year (Published)	CDD
1	Project Assessment Reform and Modernization of Tax Administration for Southeastern Europe SECO_FAD_2016	2017	FAD
2	Project Assessment Cyprus Revenue Administration (EC-financed project)_FAD_2017	2017	FAD
3*	Strengthening Tax Administration in Low-Income Asian Countries_FAD_2017	2017	FAD
4*	Strengthening Treasury Management and Fiscal Reporting in Selected SE Asian Countries_FAD_2017	2017	FAD
5	Project Assessment 8th Annual Tax Conference in Tokyo_FAD_2017	2017	FAD
6*	Project Assessment Supporting Improved Budget Planning and Fiscal Risk Management_FAD_2017	2017	FAD
7	Project Assessment Central Bank Law Reform_LEG_2017	2017	LEG
8	Project Assessment Tunisia Reinforcing Banking Supervision Capacities at the Central Bank of Tunisia_MCM_2017	2017	MCM
9	Project Assessment West Bank and Gaza Budget Preparation and Macro-fiscal Forecasting_FAD_2018	2018	FAD
10	Project Assessment Greece: Revenue Administration Project (EU-Funded)_FAD_2018	2018	FAD
11	Project Assessment JSA7: Strengthening Tax Administration in Low-Income Asian Countries_FAD_2018	2018	FAD
12	Project Assessment 9th Annual Tax Conference in Tokyo_FAD_2018	2018	FAD
13	Project Assessment Strengthening Treasury Management and Fiscal Reporting in Selected SE Asian Countries (JSA6)_FAD_2018	2018	FAD
14*	Project Assessment Bolivia AMLCFT_LEG_2018	2018	LEG
15	Project Assessment Ukraine Garnishment of Bank Accounts (FO)_LEG_2018	2018	LEG
16	Project Assessment Jordan Legislation AML/CFT Supervision and FIU TA_LEG_2018	2018	LEG
17	Project Assessment Sudan: Financial Stability_MCM_2018	2018	MCM
18	Project Assessment Democratic Republic of the Congo: Monetary Policy Advisor_MCM_2018	2018	MCM
19	Project Assessment Modernizing Central Bank Operations in Turkmenistan_MCM_2018	2018	MCM
20	Project Assessment Monetary Policy Advisor to the Governor of the Bank of Albania_MCM_2018	2018	MCM
21	Project Assessment Reinforcing Banking Supervision Capacities at the Central Bank of Tunisia_MCM_2018	2018	MCM
22	Project Assessment Suriname Central Bank Modernization_MCM_2018	2018	MCM
23	Project Assessment Sierra Leone: Central Banking Operations_MCM_2018	2018	MCM
24	Project Assessment Rwanda Financial Sector Supervision_MCM_2018	2018	MCM
25	Project Assessment Enhanced Financial Access Survey_STA_2018	2018	STA
	Project Assessment Nepal Strengthening Banking Supervision_MCM_2019	2019	MCM
26	Project Assessment Developing Macroeconomic Management Capacity in the CLMV Countries_ICD_2019	2019	ICD
27	Supporting Improved Budget Planning and Fiscal Risk Management_FAD_2019	2019	FAD
28	Project Assessment Ukraine TA on Insolvency and Creditors Rights_LEG_2019	2019	LEG
29	Project Assessment Saudi Arabia Enhancing the effectiveness of the AML/CFT framework_LEG_2019	2019	LEG
30	Project Assessment Saudi Arabia Regulatory, Supervisory and Oversight Framework for CCPs and CSDs_MCM_2019	2019	MCM
31	Project Assessment Latvia TA Project on Insolvency Evaluation_LEG_2019	2019	LEG
32	Project Assessment Southeastern Europe Strengthening Economic Governance and Public Financial Management_FAD_2019	2019	FAD

Number	Title	Year (Published)	CDD
33	Project Assessment Kyrgyz Republic Monetary Policy Frameworks_MCM_2019	2019	MCM
34	Project Assessment Fostering Financial Stability in India_MCM_2019	2019	MCM
35	Project Assessment Strengthening Liquidity Management - Bank of Central African States (BEAC)_MCM_2019	2019	MCM
36	Project Assessment Capacity Building at the National Bank of Ukraine in Selected Areas_MCM_2019	2019	MCM
37	Project Assessment Regional Government Financial Statistics_STA_2019	2019	STA
38	Project Assessment Assist the Implementation of the eGDDS (enhanced General Data Dissemination System) in Southeastern Europe_STA_2019	2019	STA
39	Project Assessment Guinea Improving Income Tax (RMTF)_FAD_2020	2020	FAD
40	Project Assessment Liberia Enhancing the AMLCFT Regime_LEG_2020	2020	LEG
41	Project Assessment National Accounts Statistics Program for Barbados_STA_2020	2020	STA
42	Project Assessment Colombia Revenue Administration Project_FAD_2020	2020	FAD
43	Project Assessment Saudi Arabia Support Establishment of a Macro-Fiscal Unit (MFU)_FAD_2020	2020	FAD
44	Project Assessment Ukraine AML to Support Anti-Corruption Effects_LEG_2020	2020	LEG
45	Project Assessment Ukraine TA on Public Financial Management (FFL)_LEG_2019	2020	LEG
46	Project Assessment Mongolia - Enhancing the AML/CFT Regime_LEG_2020	2020	LEG
47	Project Assessment Kenya Improving the AML CFT Regime - Phase II - FIU/Supervisory S&T_LEG_2020	2020	LEG
48	Project Assessment Building Capacity in Monetary Policy Modeling in Belarus_MCM_2020	2020	MCM
49	Project Assessment Nigeria Resident Advisor in Banking Supervision and Regulation_MCM_2020	2020	MCM
50	Project Assessment Bosnia and Herzegovina: Strengthening Banking Supervision_MCM_2020	2020	MCM
51	Project Assessment Bhutan: Accounting - Transitioning to BAS/IFRS_MCM_2020	2020	MCM
52	Project Assessment ECCB Strengthening Debt Management: FY14-21_MCM_2020	2020	MCM
53	Project Assessment Djibouti Islamic Banking Supervision and Regulation_MCM_2020	2020	MCM
54	Project Assessment Montenegro - IMF02 - Financial Supervision (FY17-19)_MCM_2020	2020	MCM
55	Project Assessment Strengthening Financial Systems in the ECCU_MCM_2020	2020	MCM
56	Project Assessment Capacity Building at the Ministry of Finance of Ukraine_MCM_2020	2020	MCM
57	Project Assessment Strengthening Monetary Policy Framework in the Bank of Uganda_MCM_2020	2020	MCM
58	Project Assessment Sierra Leone Strengthening Banking Supervision_MCM_2020	2020	MCM
59	Project Assessment Central Asia and South Caucasus Fiscal Transparency (GO)_STA_2020	2020	STA
60	Project Assessment Improve Monetary and Financial Statistics, Financial Soundness Indicators, and External Sector Statistics in Ukraine and North Macedonia_STA_2020	2020	STA
61	Project Assessment Improve Monetary and Financial Statistics and Financial Soundness Indicators in Eastern and Southeastern Europe_STA_2020	2020	STA

Source: IMF documents.

ANNEX IV. INTERNAL EVALUATIONS UNDERTAKEN BY CDDs DURING 2011–2020

Number	Title	Year (Published)	CDD
1	*Peru Report on the Statistics Technical Assistance and Training Evaluation Mission	2012	STA
2	External Evaluation of FAD's Advice on Government Accounting	2012	FAD
3	Valuation of MCM Technical Assistance in Belize, Costa Rica and Panama	2012	MCM
4	*Review of FAD Technical Assistance Across Three East African Countries	2013	FAD
5	Midterm Evaluation Report on the Enhanced Data Dissemination Initiative Project	2013	STA
6	Macedonia Report on the Statistics Technical Assistance and Training Evaluation Mission	2013	STA
7	Bhutan Report on the Statistics Technical Assistance and Training Evaluation Mission	2013	STA
8	*A Selective Review of Ongoing Projects of the Tax Policy and Administration Topical Trust Fund (TPA-TTF) FAD (and others)	2013	FAD
9	Bangladesh, Maldives, Nepal Evaluation report of the MCM Technical Assistance Project on Banking Supervision	2014	MCM
10	Swaziland Technical Assistance Evaluation Financial Sector Supervision and Regulation	2014	MCM
11	IMF engagement with countries in post conflict and fragile situations stocktaking	2015	SPR
12	*Guatemala Technical Assistance Report-Statistics Technical Assistance and Training Evaluation Mission	2015	STA
13	Strengthening Government Financial Integrity in Africa	2016	FAD
14	The Gambia Strengthening Budget Formulation within a Medium-Term Framework	2016	FAD
15	Evaluation of Technical Assistance on Inflation Targeting in Emerging Eastern European and Central European countries	2016	MCM
16	Building fiscal capacity in fragile states	2017	FAD
17	*Evaluation of FAD Advice on Tax Policy and Revenue Administration in Africa	2017	FAD
18	Zambia Evaluation of MCM Multi-topic Technical Assistance on Central Banking (Delivered in corporation with Norway)	2017	MCM
19	*Malawi Evaluation of MCM multi-topic Technical Assistance on Central Banking (Delivered in cooperation with Norway)	2017	MCM
20	*Evaluation of FAD Advice on Treasury Single Account and Cash management in selected Sub-Saharan African countries	2019	FAD
21	*Africa Regional Debt Management Evaluation Project	2019	MCM
22	Taking stock of IMF Capacity Development on Monetary Policy Forecasting and Policy Analysis Systems	2021	MCM

Source: IMF documents.
Note: Reviewed evaluations are marked "**".

ANNEX V. EXTERNAL EVALUATIONS OF IMF OR RTAC CD, 2012–2021

Number	Title	Year (Published)
1	The Anti-Money Laundering and Combating the Financing of Terrorism Topical Trust Fund Independent Evaluation Volume I Final Report 2012	2012
2	East AFRITAC - Independent Mid-Term Evaluation Phase III: October 2009 to date Volume I: Final Report and Volume II: Case Studies and Desktop Reviews	2013
3	West AFRITAC - Independent Mid-Term Evaluation Phase III: October 2009 to date Volume I: Final Report, Volume II: Case Studies and Desktop Reviews	2013
4	CAPTAC-DR - Independent Mid-Term Evaluation Phase I: July 2009 to date Volume I: Final Report, Volume II: : Case Studies and Desktop Reviews	2013
5	The Financial Sector Reform and Strengthening Initiative (FIRST): Phase II Evaluation	2014
6	METAC-Mid-Term Evaluation Independent External Evaluation	2014
7	Making the Transition to a Program-Based Approach to Technical Assistance (funded by JSA)	2014
8	IMF Tax Policy and Administration Topical Trust Fund TTF Independent Mid-Term Evaluation	2015
9	Independent External Evaluation of the Switzerland Technical Assistance Sub-Account of the IMF (Letters of Understanding for the East and South Work Programmes) Final Evaluation Report	2015
10	AFRITAC South Independent Mid-Term Evaluation Phase 1 June 2011 to April 2014 Volume I Final Report	2015
11	Mid-Term External Evaluation of the Managing Natural Resource Wealth Topical Trust Fund Final Report Part 1 and 2	2015
12	PFTAC Independent Mid-Term External Evaluation Report	2015
13	CARTAC-Independent External Mid-Term Evaluation Phase IV	2015
14	Central AFRITAC Independent Mid-Term Evaluation Phase II April 2011 to April 2014 Volume I Final Report	2016
15	Evaluation of the Japan Technical Assistance Subaccount (JSA)_Final Report_Volume1-Summary Report	2017
16	ATI External Evaluation	2017
17	Mid-Term Evaluation of FIRST Phase III Operations (2013-2017), Volume I: Final Synthesis Report	2017
18	*TADAT Mid-Term Evaluation_2018	2018
19	*External Evaluation of the Debt Management Facility (DMF), Phase II, Final Draft Report- Volume I and Volume II	2018
20	*AFRITAC West 2-External Mid-Term Evaluation Final Report and Annexes to the Final Report	2018
21	*CAPTAC-DR External Mid Term Evaluation Revised Final Report and Annexes	2018
22	*AFRITAC East Phase IV Midterm Evaluation Part 1 and 2	2018
23	*Independent Evaluation of Second Phase of the Anti-Money Laundering/Combating the Financing of Terrorism Thematic Trust Fund	2019
24	*Independent Mid-Term Evaluation of the Somalia Country Fund Final Evaluation Report	2019
25	*Independent Evaluation Switzerland Technical Assistance Subaccount State Secretariat for Economic Affairs (SECO)_ 2020	2020
26	*TADAT Phase 1 Completion Report, 2020	2020
27	*METAC- Phase IV Independent Mid-Term Evaluation	2021

Source: IMF documents.
Note: Reviewed evaluations are marked "**".

ANNEX VI. WHAT FACTORS CORRELATE WITH CD OUTCOME RATINGS?

AN APPLICATION OF BASSANETTI (2021) FRAMEWORK

By Luigi Durand

INTRODUCTION

The significance of Capacity Development as a major driver of IMF spending motivates further analysis of the processes and systems at its core. In this respect, the 2018 CD Strategy Review raised specific recommendations, including the importance of strengthening the coordination, communication and dissemination of good practices, the relevance of fine-tuning the delivery, and the necessity of enhancing the monitoring of activities and narrowing the list of priorities (IMF, 2018b).

Within this context, the Results-Based Management (RBM) approach, by operationalizing CD activities' outcomes allows for a rigorous assessment that can help inform policymakers on what drives CD successes. In particular, the emphasis on outcomes that are rated based on the achievement of pre-specified targets, lends itself to analysis that can quantify CD progress, and ultimately also inform on the role of potential correlates of interest.¹

To this end, Bassanetti (2021) investigates the role of country-specific macroeconomic conditions and CD specific project features in influencing CD outcome ratings, using a dataset of approximately 3000 RBM outcome ratings from IMF CD projects started between 2013 and 2019 for a total of 152 recipient countries. Bassanetti finds a statistically significant positive association between outcome ratings and the growth rate of GDP, the presence of a resident advisor and short-term experts on the ground, the involvement of an RTAC, the degree of project completion and the concomitant presence (and completion) of an IMF program in the country. Inversely, outcome ratings are found to be significantly lower in the case of projects with many workstreams and in the case of projects carried out in Fragile States.

This note reports the findings of an IEO exercise to replicate Bassanetti's approach and examine the sensitivity of his results to changes in data sample. We do this because the amount of RBM data held by the Fund on CD projects is increasingly rapidly and outcome ratings are updated periodically through the lifetime of CD projects, which could make a difference to the findings of B's analysis. Our baseline sample covers the same period (i.e., CD projects started between 2013 and 2019) but includes around 45 percent more rated outcomes, which were added to the RBM database between February 2020 (when Bassanetti extracted his sample) and December 2020, when the data for this analysis was extracted. Given the potentially important effects of the pandemic, we also investigate whether the results are affected by some project outcome ratings having been added or amended since March 2020.

¹ Outcome Ratings are based on the status of indicators and milestones set at the outset of the project and are instrumental to quantify the achievement of CD objectives.

Model Specification

Bassanetti starts with an ordered Probit model describing the relationship between outcomes ratings and a set of variables of interest. More specifically,

$$\Pr(OR_{kijt} = z \mid CF_{jt}, PF_{ijt}, CO_{kijt}) = \Phi(\beta CF_{jt} + \gamma PF_{ijt} + \omega CO_{kijt}) + \epsilon_{kijt} \quad \textbf{(Equation 1)}$$

with $z = 1, 2, 3, 4$ and where OR_{kijt} refers to the outcome k rating, associated with project i , delivered in country j and started at time t , CF refers to the set of macroeconomic and country-level variables of interest, PF refers to a set of project features and CO is a set of control variables.

The macroeconomic variables of interest are the level, growth rate and volatility of per-capita real GDP, and dummy variables for fragile states, small states, and for countries having IMF programs (with a distinction between programs that have been completed and those that have gone off-track).²

The “project features” variables of interest are a dummy variable that distinguishes between outcomes associated with RTAC involvement or not (and in particular with specific dummies for African RTACs, Central American and Caribbean RTACs, and the rest³), the amount of resources deployed in terms of resident advisors, STX and HQ personnel (all quantified in terms of FTEs), the number of project workstreams (as a proxy for project complexity), and a measure of how far the project is through its intended duration (“project completion share”).

The control variables of interest are a set of dummy variables for the initial year of the project, the year when the outcome was last rated, the geographical region and income group of the country associated with a project and the workstream associated with the outcome. Compared to Bassanetti, the model also features a “COVID-19” dummy equal to 1 for all those outcome ratings that have been updated on or after March 2020.

Appendix to Annex VI defines the variables used in the regression analysis and presents the main results.

Data Sample and Summary Statistics

Bassanetti’s sample was drawn from the RBM database as it stood in February 2020, covering CD projects composing the dataset started between 2013 and 2019. His sample included 2957 rated outcomes involving 152 recipient countries.

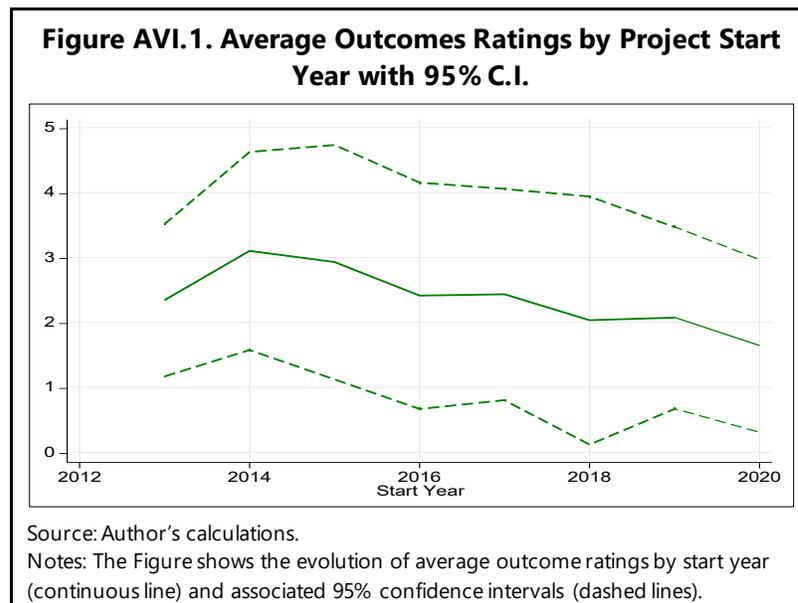
² We use the same dataset of Bassanetti to categorize IMF programs. One important caveat is that the available information on programs stops in 2018, which means that our regressor misses all programs starting after that date; it follows that the resulting estimated coefficients should be evaluated with care.

³ When defining the RTACs dummy variables the African group includes AFRITAC Central, AFRITAC South, AFRITAC West 2, AFRITAC East and AFRITAC West, while the Central American / Caribbean group includes CAPTAC and CARTAC. This distinction between groups of RTACs differ from Bassanetti’s analysis where instead one single dummy includes all RTACs together.

The baseline sample used in this analysis was drawn from the RBM database as it stood in [December 2020], and like Bassanetti's sample includes CD projects started between 2013 and 2019. This baseline sample thus excludes all projects that started after January 2020, in order to limit the impact of a potential structural break in the data associated with the pandemic.⁴ The baseline sample for this analysis consists of 4278 rated outcomes, some 45 percent larger than Bassanetti's sample.

Figure AVI.1 summarizes the average outcome ratings in RBM by CD project start year (including the ones associated with projects that started in 2020); the data reveals a decline in ratings associated with projects started more recently. The decline is (at least partially) expected, since outcome ratings are updated annually, until the project is completed; and typically, as a project progresses, outcome ratings improve. These dynamics are further suggested by Table AVI.1, where the ratio of approved/completed projects in the sample is naturally increasing as we move to more recent start years.

Figure AVI.2 summarizes the number of observations (i.e., outcome ratings) organized by project start year and status of the associated project. The majority of outcome ratings in the sample are from ongoing projects. The large number of projects in the sample notionally started in 2017 likely reflects a push within the IMF to extend the coverage of the RBM approach.

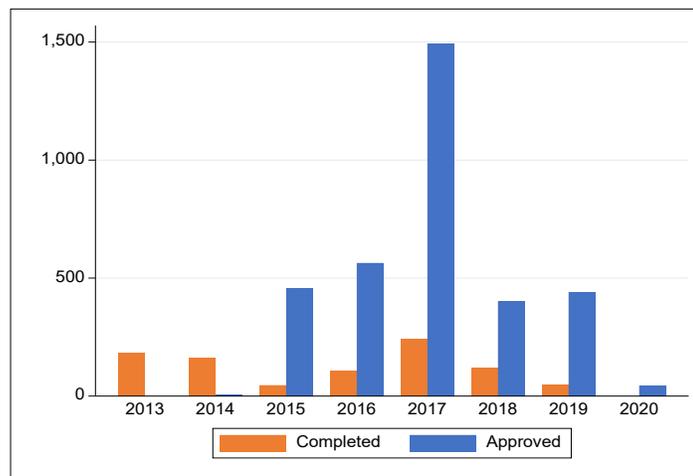


⁴ In practice, this means removing only 36 outcome ratings, from a full sample of 4314 rated outcomes, associated with 379 projects.

Table AVI.1 Number of Completed and Approved Projects by Start Year

Start Year	Completed	Approved	Approved/Completed
2013	5	0	0.00
2014	15	2	0.13
2015	6	10	1.67
2016	13	22	1.69
2017	52	96	1.85
2018	27	70	2.59
2019	5	45	9.00
2020	0	8	-

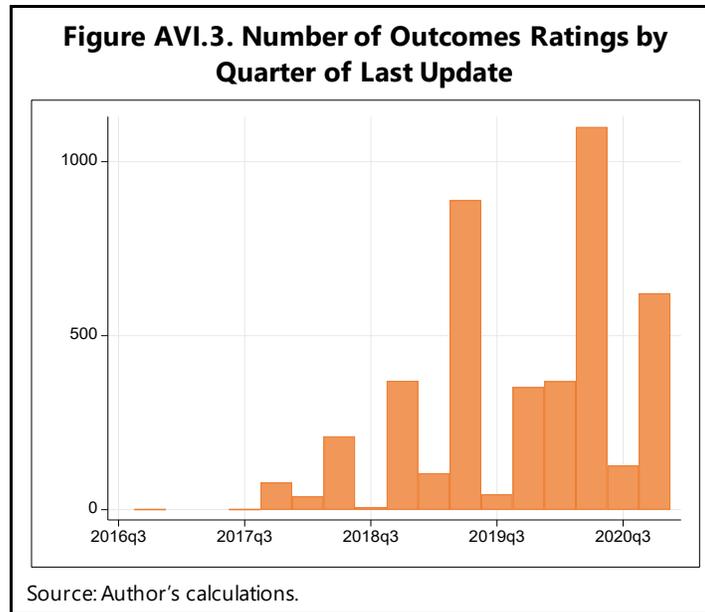
Notes: The table shows the number of completed and approved projects by start year, including projects started in 2020.

Figure AVI.2 Number of Outcomes Ratings by Project Start Year and Status

Source: Author's calculations.

Notes: The Figure shows the total number of rated outcomes by project status and start date of the project. As an example, the Figure highlights that the total number of rated outcomes associated with approved projects that were started in 2017 is approx. 1500, while the total number of outcomes associated with completed projects that were started in 2017 is approx. 250.

The sample used in this note is larger than Bassanetti's sample (4071 versus 2836 observations in total used in the regression, excluding the 36 ratings associated with 2020 projects). This is because while Bassanetti's analysis was carried out in spring 2020, this exercise exploits an updated database, which includes ratings up to December 2020. In this respect, note that 2020q2 saw a large spike of 1104 entries followed by a total of approximately 722 entries between 2020q3 and 2020q4 (Figure AVI.3).



To conclude the description of the sample, Appendix to Annex VI presents a selection of summary statistics for the outcome ratings, split by different categories of interest. The main points are as follows:

- Overall, compared to Bassanetti, the average outcome rating over the whole sample is similar (2.43 versus 2.35).⁵
- As in Bassanetti, the average outcome ratings are higher (compared to the overall average) in the case of completed projects (2.66), lower in the case of ongoing projects (2.36), and lower in the case of Fragile Countries (FCS) and Small Developing countries (SDS; 2.37 and 2.34, respectively);⁶ in Bassanetti completed projects have a 2.64 average rating, ongoing projects a 2.31 average rating, FCS a 2.22 average rating and SDS a 2.31 average rating.
- Interestingly, low-income countries tend to fare relatively better when compared to high-income countries (2.40 and 2.35, respectively), which is in contrast with Bassanetti's sample (where the means equal 2.29 and 2.44, respectively). Also, middle-income

⁵ A t-test reveals that the difference in means is statistically significant at 1 percent level.

⁶ A t-test reveals that in the case of completed projects, the difference in means is statistically insignificant while in the case of ongoing projects the difference in means is statistically significant at the 5 percent level; in the case of analytical groups, a similar test reveals statistically significant differences at the 1 percent level for the case of Fragile and Conflict-Affected Countries but statistically insignificant differences in the case of Small Developing States.

countries have an average higher rating when compared to Bassanetti's sample (2.45 versus 2.41, respectively).⁷

- Furthermore, average outcome ratings differ across geographical regions; specifically, Europe and Central Asia score the highest (2.48), followed by Latin America and Caribbean and Sub-Saharan African countries (both 2.46), while Middle East and North Africa and North American countries score the lowest (2.35 and 2.08, respectively). Bassanetti relies on a slightly different geographic category definitions and finds that the Western Hemisphere scores the highest (2.59), followed by Middle East and Central Asia (2.36), Asia-Pacific and Europe (both 2.29) and Sub-Saharan Africa (2.27).⁸

Results

Figures AVI.4 and AVI.5 summarize the main findings from the model described in Eq. (1) above. The corresponding regression results are shown in the Appendix to Annex VI.

When we use the same model specification as in Bassanetti (column 1 in Table AVI.3) the results are broadly similar; the two main differences relate to the impact of RTACs (positive and highly significant in Bassanetti's results, insignificant in our results) and the amount of LTX input to the project (positive and significant in Bassanetti, negative and significant in our results). When we add a "COVID" dummy for outcome ratings added or amended since March 2020 (column 2 in Table AVI.3), this dummy variable is significant (and marginally improves the overall fit of the model as measured by the pseudo R-squared) but doesn't change the differences with Bassanetti's results.

In column 3 we also include additional dummy variables to distinguish between different RTAC involvement. This is our preferred baseline model specification. The results suggest that, on average, RBM outcome ratings are:

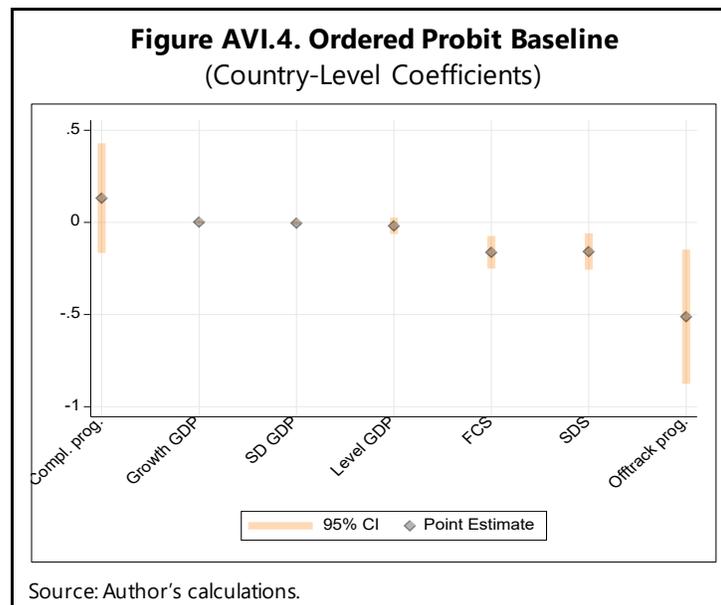
- Lower in fragile and conflict affected countries (at the 1 percent significance level).
- Higher in the case of greater deployment of resources by STX and HQ personnel (expressed in terms of FTEs); curiously, this latter result is flipped in the case of resident advisors (both results hold at the 1 percent significance level).
- Lower in the case of programs that are off-track (at the 1 percent significance level).

⁷ A t-test reveals that the differences are statistically significant at the 1 percent level for the case low-income countries but statistically insignificant in the case of High-income countries. Also, the difference is statistically insignificant in the case of middle-income countries (which are defined as "emerging markets" in B).

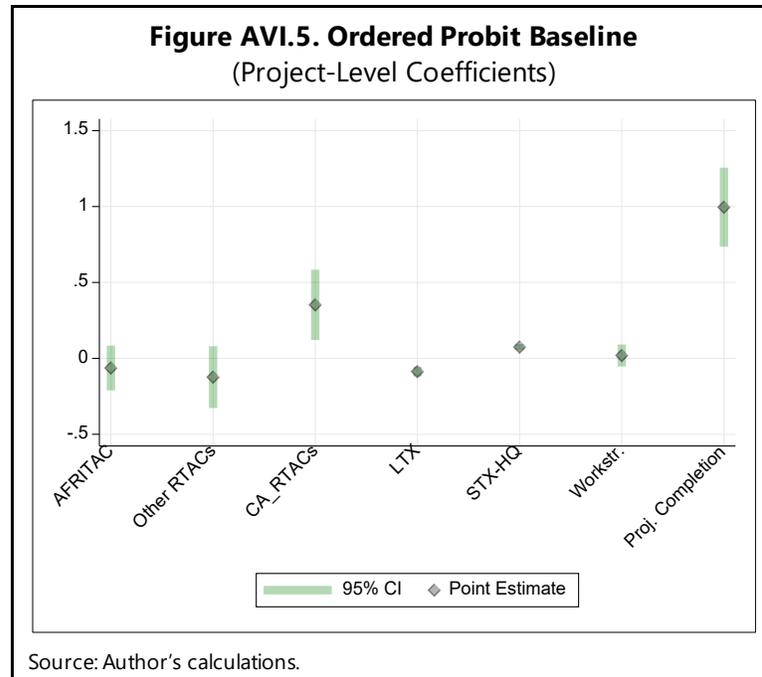
⁸ The differences in classifications used do not allow to test the differences in means, with the exception of the Sub-Saharan category (which coincides with our paper), where we find a statistically significant difference at the 1 percent level.

- Higher in the case of more mature projects, as measured by the project completion share (at the 1 percent significance level).⁹
- Higher in the case of projects carried out by central American and Caribbean RTACS ("CA1", "CA2"; at the 1 percent significance level) but lower in the case of projects carried out by African RTACs ("AFC", "AFE", "AFS", "AFW", "AW2"; albeit without statistical significance), and by other RTACs ("MTA", "PFT" and "STC"; albeit without statistical significance) compared to ratings associated to projects without RTACs.
- Lower when those outcome ratings have been added or amended during the COVID-19 crisis (at the 1 percent significance level).

The model finds no statistically significant correlation of CD project outcomes with GDP (level), GDP growth or volatility of output, or with the country having completed IMF programs.



⁹ This result in particular is expected by "design," since outcomes ratings are regularly updated as the project evolves.



While our results broadly confirm Bassanetti's findings, there are some interesting differences. In the next section we further attempt to reconcile these results with those in Bassanetti by pinning down the role that COVID-19 and certain multi-country projects have in driving the sign and significance of several key regressors of interest.

Sensitivity

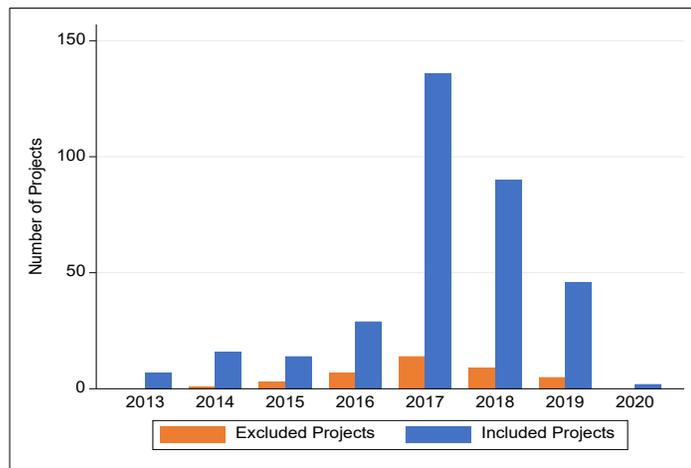
This section explores the sensitivity of the baseline results to alternative specifications and sample sizes.

First, we consider different ways to measure project complexity, as also done in Bassanetti. One possibility is to replace the number of workstreams by projects and delivery department (as done in the baseline regression) by a measure that instead calculates the number of objectives in the project that are associated with the workstream within which the relevant targeted outcome falls; a second possibility is to instead compute the number of outcomes in the project associated with the objective within which the targeted outcome falls. Table A.I.IV in the Appendix to Annex VI summarizes these results; interestingly, while the "number of workstreams" and the "number of outcomes" do not show any statistical significance, the "number of objectives" is highly significant and positive, suggesting that having more objectives within a workstream is associated with better outcome ratings on average. Interestingly, the introduction of the "number of objectives" regressors wipes out the statistically significant result found for the case of Central American RTACs. The two main differences between our results and Bassanetti (namely, different signs on the coefficients for LTX and RTACs) remain.

We then turn to two additional sensitivity analyses in relation to the sample size at hand. First, and in order to remove any potential source of structural break in the data, we perform a regression that excludes all outcome ratings that have been updated after February 2020; this results in a much smaller sample size, of 2351 compared to 4278 in our baseline analysis. Second, we run a separate regression using a sample that includes projects with outcome ratings amended or added after February 2020 but excludes all projects that have AFR, APD, EUR, IMF or MCD as a beneficiary country. These 39 projects have a total 879 outcome ratings and involve several delivery departments: STA (533 ratings), FAD (243 ratings), MCM (102 ratings) and LEG (1 ratings). The removal of these projects, resulting in a sample size of 3400 rated outcomes, is motivated by two “unusual” patterns, that cannot be easily reconciled with the rest of the dataset. In particular, the “excluded” projects tend to have a much larger number of objectives (approximately 48 objectives per project in the case of the excluded projects versus around 20 objectives per project on average in the remaining sample) and also have an underlying distribution of outcome ratings that does not seem to follow a “Gaussian behavior” (see Figure 7). In particular, while in the rest of the dataset the bulk of ratings are clustered around 2 and 3, the “excluded” projects ratings are skewed, with the majority of the projects being rated 4 and 3.

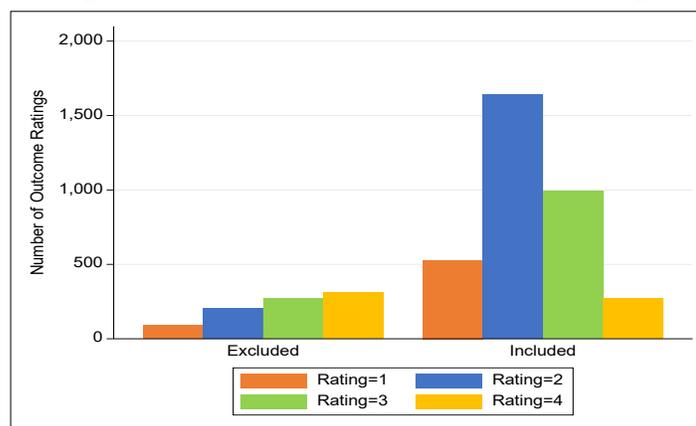
The results in Table AVI.5. summarize the regression results; the first column, which excludes the subset of ratings that were updated after February 2020, confirms that the baseline results are robust to the exclusion of the sample of data that overlaps with the COVID-19 pandemic; one major difference concerns the growth rate of GDP, which is now positively associated with the outcome ratings (at the 1 percent significance level). The second and third columns exclude the categories of projects highlighted in the previous paragraph; in this case the regression results reveal major differences; in particular, we now observe that, in line with Bassanetti’s results, the RTAC dummy becomes positive and significant (at the 1 percent level; this result also holds true when the RTAC dummy is split into regional dummies, as shown in column 3 of the Table);¹⁰ furthermore, and in contrast with the baseline results, the amount of LTX has now a positive impact on outcomes success (as per Bassanetti), while the amount of STX has a negative impact (the former at the 1 percent significance level and the latter at the 5 percent significance level), contrary to Bassanetti’s results. The coefficients on the dummy variable for SDS and the off-track dummy variable both lose statistical significance; finally, the distinct workstream dummy variable, while remaining statistically insignificant now display a negative term. Overall, these results show a much better alignment with Bassanetti’s analysis and point to the distinctive role that the “excluded” projects (i.e., those where the beneficiary is a region rather than a specific country) have in outcome ratings. Further research should focus on understanding the nature of the excluded sample of projects, with regards to both the structure of the projects themselves (number of milestones, workstreams etc.) but also the reasons behind the unusual rating distribution.

¹⁰ In particular, the coefficient is highest in the case of Central American RTACs, followed by the “other” RTACs category and the African RTACs.

Figure AVI.6. Excluded Projects by Start Year

Source: Author's calculations.

Notes: The excluded projects refer to the CD projects having AFR, APD, EUR, IMF or MCD as a beneficiary country. The included projects are the remaining projects in the sample.

Figure AVI.7. Distribution of Outcome Ratings

Notes: The Figure shows the distribution of outcome ratings in the excluded sample versus the included sample. The excluded projects refer to the CD projects having AFR, APD, EUR, IMF or MCD as a beneficiary country. The included projects are the remaining projects in the sample.

APPENDIX TO ANNEX VI. MAIN VARIABLES AND RESULTS

A.I. Variables Description

Real GDP per capita – Source: World Development Indicator (World Bank).

Growth rate of real GDP per capita – Computed as the average growth rate of per capita GDP over the 5 years preceding the start of the CD project. Source: World Development Indicator (World Bank).

Volatility of real GDP per capita – Computed over the 10 years preceding the start of the CD project. Source: World Development Indicator (World Bank).

Fragile and Conflict Affected States list – Source: “The IMF and Fragile States”, IEO Evaluation Report, 2018.

Small Developing States list – Source: “IMF Engagement in Small States,” IEO Draft Issues Paper, 2020.

FTEs of LTX, STX, HQ – For each CD project and objective country, FTEs are summed, by type of delivery mechanism (LTX, STX, HQ). Source: TIMS Dataset.

RTAC dummy variable – Dummy variable equal to 1 in case of a project involving an RTACs. Different dummies are applied to different regions (see main text for details). Source: RBM and ICD calculations.

Program Status – The model includes both a dummy variable for “completed” programs and also a dummy variables for programs that are “off-track.” Source: SPR dataset on programs.

Project Completion Share: This variable is constructed as:

$$\min\left\{\frac{\text{last updated date for RBM outcomes Ratings} - \text{project start date}}{\text{project end date} - \text{project start date}}, 1\right\}$$

Source: RBM.

Geographical Regions – Source: World Bank Classification.

Income Groups – Source: World Bank Classification.

Number of Workstreams by project – Source: RBM.

Covid Dummy – Variable equal to one if the outcome ratings last update is made on or after March 2020, and zero otherwise.

Table A.I.1. Summary Statistics from Bassanetti (2021)

Dataset Composition	Number of Rated Outcomes	Original Ratings (1–4)		Binary Ratings (0–1)	
		Mean	Standard Deviation	Mean	Standard Deviation
<i>Total dataset</i>	2957	2.35	0.81	0.37	0.48
<i>Of which:</i>					
Completed projects	321	2.64	0.90	0.51	0.50
Ongoing projects	2636	2.31	0.79	0.36	0.48
<i>Income groups</i>					
Advanced economies	27	2.44	0.97	0.41	0.50
Emerging markets	1317	2.41	0.89	0.42	0.49
Low-income countries	1613	2.29	0.73	0.34	0.47
<i>Analytical groups</i>					
Fragile and conflict affected states	1043	2.22	0.71	0.28	0.45
Small Developing States	537	2.31	0.75	0.37	0.48
<i>Macro-regions</i>					
Asia-Pacific	431	2.29	0.84	0.35	0.48
Europe	241	2.29	0.97	0.32	0.47
Middle East and Central Asia	311	2.36	0.83	0.38	0.49
Sub-Saharan Africa	1381	2.27	0.69	0.32	0.47
Western Hemisphere	593	2.59	0.91	0.54	0.50

Source: IEO staff analysis.

Notes: As of February 2020. Author's calculation.

Table A.I.2. Summary Statistics

	Number of Observations	Mean	Standard Deviation
<i>Number of rated outcomes of which:</i>	4278	2.43	0.89
completed projects	907	2.66	0.96
ongoing projects	3,371	2.36	0.86
<i>Income groups</i>			
High Income	503	2.35	0.87
Low Income	1,114	2.40	0.83
Middle Income	2,661	2.45	0.92
<i>Analytical Groups</i>			
Fragile and conflict affected states	1,192	2.37	0.85
Small Developing States	825	2.34	0.91
<i>Geographic Regions</i>			
East Asia & Pacific	542	2.26	0.92
Europe & Central Asia	475	2.48	0.98
Latin America & Caribbean	891	2.46	0.87
Middle East & North Africa	221	2.35	1.01
North America	13	2.08	0.76
South Asia	158	2.36	1.05
Sub-Saharan Africa	1,978	2.46	0.84

Source: IEO staff analysis.

Notes: This table shows the summary for the sample associated with the baseline regression in the text.

Table A.I.3. Baseline Regression and Comparison with Bassanetti (2021)

Dependent Variable Outcome Ratings (1–4)	(1)	(2)	(3)	Bassanetti (2021)
Log of per capita real GDP	-0.021 (0.023)	-0.015 (0.023)	-0.019 (0.023)	-0.0055 (0.0407)
Per capita real GDP growth	0 (0.004)	0.001 (0.004)	0.002 (0.004)	0.0402*** (0.0085)
Volatility of per capita real GDP growth	-0.0050 (0.004)	-0.0040 (0.004)	-0.0040 (0.004)	-0.3230 (0.6258)
FCS Countries dummy	-0.161*** (0.045)	-0.161*** (0.045)	-0.162*** (0.045)	-0.1023* (0.0604)
SDS Countries dummy	-0.131*** (0.050)	-0.133*** (0.050)	-0.158*** (0.050)	-0.1324* (0.0678)
FTE of LTX	-0.074*** (0.016)	-0.082*** (0.016)	-0.087*** (0.016)	0.1053*** (0.0384)
FTE of STX and HQ	0.060*** (0.010)	0.073*** (0.010)	0.075*** (0.010)	0.0864** (0.0421)
All RTACs dummy	-0.038 (0.061)	-0.021 (0.061)	-	0.4552*** (0.0756)
African RTACs dummy	-	-	-0.063 (0.075)	-
Central American RTACs dummy	-	-	0.352*** (0.118)	-
Other RTACs dummy	-	-	-0.123 (0.104)	-
IMF off-track program	-0.455** (0.184)	-0.487*** (0.187)	-0.511*** (0.186)	-0.1264 (0.0921)
IMF Completed Program	0.206 (0.149)	0.169 (0.150)	0.132 (0.152)	0.1475** (0.0675)
Project Completed share	0.402*** (0.099)	0.981*** (0.132)	0.996*** (0.133)	0.0135*** (0.0018)
Number of Workstream	0.031 (0.036)	0.027 (0.036)	0.019 (0.037)	-0.3554*** (0.0734)
COVID dummy	-	-0.419*** (0.055)	-0.438*** (0.056)	-
Observations	4278	4278	4278	4278
Pseudo R2	0.096	0.1021	0.1036	0.0698
Project start year dummies	Yes	Yes	Yes	Yes
Project evaluation year dummies	Yes	Yes	Yes	Yes
Workstream dummies	Yes	Yes	Yes	Yes
Macro-region dummies	Yes	Yes	Yes	Yes
Income group dummies	Yes	Yes	Yes	Yes

Source: IEO staff analysis.

Notes: The table reports the estimated coefficients and, in brackets, the associated robust standard errors. ***, ** and *, indicate significance at the 1, 5, 10 percent, respectively.

Column (1) reports the results without COVID-19 dummy, column (2) introduces the COVID-19 dummy and column (3) disaggregate the RTAC variables; finally, column Bassanetti (2021) refers to the results presented in the original IMF WP.

Table A.I.4. Sensitivity Analysis – Alternatives Definitions of Project Complexity

Dependent Variable Outcome Ratings (1–4)	(1)	(2)	(3)
Log of per capita real GDP	-0.019 (0.023)	-0.02 (0.023)	-0.019 (0.023)
Per capita real GDP growth	0.002 (0.004)	0.002 (0.004)	0.002 (0.004)
Volatility of per capita real GDP growth	-0.0040 (0.004)	-0.0040 (0.004)	-0.0040 (0.004)
FCS Countries dummy	-0.162*** (0.045)	-0.186** (0.045)	-0.164*** (0.045)
SDS Countries dummy	-0.158*** (0.050)	-0.170*** (0.050)	-0.158*** (0.050)
FTE of LTX	-0.087*** (0.016)	-0.068*** (0.017)	-0.088*** (0.016)
FTE of STX and HQ	0.075*** (0.010)	0.040*** (0.011)	0.078*** (0.009)
All RTACs dummy	-	-	-
African RTACs dummy	-0.063 (0.075)	-0.053 (0.071)	-0.079 (0.071)
Central American RTACs dummy	0.352*** (0.118)	0.184 (0.125)	0.346*** (0.119)
Other RTACs dummy	-0.123 (0.104)	-0.113 (0.104)	-0.126 (0.104)
IMF off-track program	-0.511*** (0.186)	-0.566*** (0.189)	-0.506*** (0.185)
IMF Completed Program	0.132 (0.152)	0.12 (0.153)	0.135 (0.152)
Project Completed share	0.996*** (0.133)	1.002*** (0.133)	0.994*** (0.133)
Number of Workstream	0.019 (0.037)	-	-
Number of Objectives	-	0.007*** (0.001)	-
Number of Outcomes	-	-	-0.003 (0.014)
COVID dummy	-0.438*** (0.056)	-0.434*** (0.056)	-0.438*** (0.056)
Observations	4278	4278	4278
Pseudo R2	0.1036	0.1057	0.1036
Project start year dummies	Yes	Yes	Yes
Project evaluation year dummies	Yes	Yes	Yes
Workstream dummies	Yes	Yes	Yes
Macro-region dummies	Yes	Yes	Yes
Income group dummies	Yes	Yes	Yes

Source: IEO staff analysis.

Notes: The table reports the estimated coefficients and, in brackets, the associated robust standard errors. ***, ** and *, indicate significance at the 1 percent, 5 percent, and 10 percent, respectively.

Column (1) reports the results with project complexity defined in terms of number of workstreams; column (2) reports the results with project complexity defined in terms of number of objectives and column (3) reports the results with project complexity defined in terms of number of outcomes. Refer to the text for more details.

Table A.I.5. Sensitivity Analysis – Alternatives Samples

Dependent Variable Outcome Ratings (1–4)	(1)	(2)	(3)
Log of per capita real GDP	-0.032 (0.033)	0.002 (0.029)	0.002 (0.029)
Per capita real GDP growth	0.018*** (0.006)	0.003 (0.005)	0.004 (0.005)
Volatility of per capita real GDP growth	-0.0020 (0.006)	0.0060 (0.004)	0.0060 (0.004)
FCS Countries dummy	-0.161*** (0.062)	-0.186*** (0.055)	-0.183*** (0.055)
SDS Countries dummy	-0.136** (0.069)	-0.013 (0.055)	-0.027 (0.055)
FTE of LTX	-0.027 (0.023)	0.074*** (0.021)	0.072*** (0.022)
FTE of STX and HQ	0.021 (0.015)	-0.030** (0.014)	-0.028** (0.014)
All RTACs dummy	0.132 (0.090)	0.259*** (0.077)	-
African RTACs dummy	-	-	0.196** (0.089)
Central American RTACs dummy	-	-	0.391*** (0.134)
Other RTACs dummy	-	-	0.305** (0.143)
IMF off-track program	-0.519** (0.223)	-0.374 (0.232)	-0.365 (0.233)
IMF Completed Program	0.218 (0.179)	0.177 (0.162)	0.176 (0.165)
Project Completed share	1.266*** (0.171)	0.920*** (0.151)	0.917*** (0.152)
Number of Workstream	0.169** (0.070)	-0.018 (0.050)	-0.024 (0.050)
COVID dummy	-	-0.447*** (0.061)	-0.457*** (0.062)
Observations	2351	3400	3400
Pseudo R2	0.068	0.065	0.066
Project start year dummies	Yes	Yes	Yes
Project evaluation year dummies	Yes	Yes	Yes
Workstream dummies	Yes	Yes	Yes
Macro-region dummies	Yes	Yes	Yes
Income group dummies	Yes	Yes	Yes

Source: IEO staff analysis.

Notes: The table reports the estimated coefficients and, in brackets, the associated robust standard errors. ***, ** and *, indicate significance at the 1, 5, 10 percent, respectively. Column (1) reports the results from a model that excludes all outcomes ratings that are updated after February 2020, column (2) reports the results from a model that excludes certain multi-country projects, and column (3) reports the results from a model that excludes the ratings associated with certain multi-country projects and split the RTAC dummy by regions. Refer to the text for more details.

REFERENCES

- Asian Development Bank, 2014, "Role of Technical Assistance in ADB Operations" Corporate Evaluation Study, September (Manila).
- Bassanetti, Antonio, 2021, "When Does Capacity Development Achieve Good Outcomes? Evidence from the IMF Results-Based Management Data," IMF Working Paper No. WP/21/285, December (Washington: International Monetary Fund).
- Chopra, Ajai, 2022, "The IMF and Capacity Development—Case Studies for Middle East and Central Asia" IEO Background Paper No. BP/22-02/04 (Washington: International Monetary Fund).
- Chami, Ralph and Darkey, Elorm and Williams, Oral H., 2021, "A Time to Build: Does TA Matter for Revenue Mobilization?" IMF Working Paper No. 2021/035 (Washington: International Monetary Fund).
- Citrin, Daniel, and Christopher Legg, 2022, "The IMF and Capacity Development—Case Studies for Asia," IEO Background Paper BP/22-02/02 (Washington: International Monetary Fund).
- De Lannoy, Anthony, 2022, "The IMF and Capacity Development—Training," IEO Background Paper No: BP/22-02/09 (Washington: International Monetary Fund).
- Durand, Luigi, 2022, "What Factors Correlate with CD Outcome Ratings? An Application of Bassanetti (2021) Framework," reproduced in Annex VI.
- Edison, Hali, Dimitre Milkov, and Josephat Koima, 2018, "Evaluating IMF Training: What Can We Learn?" IMF Working Paper No. WP/18/151, June (Washington: International Monetary Fund).
- Enoch, Charles, 2022 "The IMF and Capacity Development—Delivery," IEO Background Paper No: BP/22-02/08 (Washington: International Monetary Fund).
- Everaert, Luc, 2022, "The IMF and Capacity Development—Case Studies for Europe," IEO Background Paper No: BP/22-02/03 (Washington: International Monetary Fund).
- Gupta, Sanjeev, and others, 2013, "Review of FAD Technical Assistance across three East African Countries" (Washington: International Monetary Fund).
- Independent Evaluation Office of the International Monetary Fund (IEO), 2005, *IMF Technical Assistance* (Washington: International Monetary Fund).
- _____, 2014, *IMF Technical Assistance: Revisiting the 2005 IEO Evaluation* (Washington: International Monetary Fund).

- _____, 2015, *Self-Evaluation at the IMF: An IEO Assessment* (Washington: International Monetary Fund).
- Independent Evaluation Group of the World Bank Group (IEG), 2021, "Results and Performance of the World Bank Group 2021" (Washington: World Bank Group).
- International Monetary Fund, 2002, "Review of Technical Assistance Policy and Experience," June (Washington).
- _____, 2008, "Enhancing the Impact of Fund Technical Assistance," Policy Paper, April (Washington).
- _____, 2011, "Report of the Task Force on the Fund's Technical Assistance Strategy," IMF Policy Paper, October (Washington).
- _____, 2013a, "The Fund's Capacity Development Strategy—Better Policies Through Stronger Institutions," IMF Policy Paper, May (Washington).
- _____, 2013b, "The Fund's Capacity Development Strategy—Better Policies Through Stronger Institutions—Background Paper" IMF Policy Paper, May (Washington).
- _____, 2014, "IMF Policies and Practices on Capacity Development," IMF Policy Paper, August (Washington).
- _____, 2015a, "Regional Technical Assistance Centers—The RTAC Handbook" (internal).
- _____, 2015b, "CD-PORT Manual" (Washington).
- _____, 2017, "New Common Evaluation Framework for IMF Capacity Development," IMF Policy Paper, April (Washington).
- _____, 2018a, "OIA Advisory Review of Results-Based Management Initiative for Capacity Development," March (internal).
- _____, 2018b, "2018 Review of the Fund's Capacity Development Strategy," IMF Policy Paper, October (Washington).
- _____, 2018c, "2018 Review of the Fund's Capacity Development Strategy—Staff Background Studies and Short Notes," October (Washington).
- _____, 2018d, "2018 Review of the Fund's Capacity Development Strategy—External Commentary and Background Studies," October (Washington).
- _____, 2018e, "2018 Review of the Fund's Capacity Development Strategy—Survey of Stakeholders and Independent Interviews," October (Washington).

- _____, 2019, "IMF Policies and Practices on Capacity Development," IMF Policy Paper, November (Washington).
- _____, 2020a, "Staff Guidelines on Capacity Development Country Strategy Notes," January (internal).
- _____, 2020b, "Institute for Capacity Development—Capacity Development Strategy," February (internal).
- _____, 2020c, "CDMAP User Manual," August (internal).
- _____, 2020d, "RBM Governance Framework," August (Washington).
- _____, 2020e, "Updated Common Evaluation Framework for IMF Capacity Development and Guidance Note," September (Washington).
- _____, 2020f, "Results Based Management (RBM) Catalog," November (internal).
- _____, 2021a, "Operational guidelines for integrating CD with Surveillance and Lending," April (internal).
- _____, 2021b, "Operational Guidelines – 2020 RBM Governance Framework," May (internal).
- _____, 2021c, "CD Evaluations and Impact," Informal Board Meeting, August 31, 2021 (internal).
- Keijzer, Niels, and Stephan Klingebiel, 2019, "A new old principle: Ownership for sustainable development cooperation," *The Current Column*, January (German Development Institute, Bonn).
- Kirkpatrick, Donald L., 1994, "Evaluating Training Programs," November, Berrett-Koehler Publishers.
- Lawson, Andrew, Gonzalo Alvarez, Bruno Giussani, Luis Carlos Jemio and Gonzalo Contreras, 2018, "External Mid-Term Evaluation of the Central America, Panama and the Dominican Republic Technical Assistance Centre (CAPTAC-DR)," FISCUS Public Finance Consultants (August).
- Legg, Christopher, and Daouda Sembene, 2022, "The IMF and Capacity Development—Case Studies for Africa", IEO Background Paper No. BP/22-02/01 (Washington: International Monetary Fund).
- Morra Imas, Linda G., and Ray C. Rist, 2009, "The Road to Results" (Washington: The World Bank).
- OECD, 2018, "Learning from Results-Based Management evaluations and reviews," Discussion paper for the OECD/DAC Results Community Workshop, October 29-30 (Paris).

- _____, 2019a, "Learning from Results-Based Management evaluations and reviews," Working Paper, March (Paris).
- _____, 2019b, Revised OECD-DAC Evaluation Criteria.
- _____, 2019c, "Managing for Sustainable Results – OECD DAC Guiding Principles," July (Paris).
- Pastor, Gonzalo, 2009, "A Framework to Assess the Effectiveness of IMF Technical Assistance in National Accounts," IMF Working Paper No. WP/09/269, (Washington: International Monetary Fund).
- Pedraglio, Roxana, and Louellen Stedman, 2022, "The IMF and Capacity Development—IEO Surveys and Analysis of AidData Surveys," IEO Background Paper No. BP/22-02/15 (Washington: International Monetary Fund).
- Radelet, Steven, 2018, "2018 Quinquennial Review of the Fund's Capacity Development Strategy—External Study—A Review of the IMF's Efforts to Strengthen Individual and Institutional Capacity" (Washington: International Monetary Fund).
- ROI Institute, 2019, "IMF Post Course Follow-up surveys, 2017-2018 Course Follow-Up," (Alabama).
- Selowski, Marcelo, and Ling Hui Tan, 2015, "Self-Evaluation in the Areas of Technical Assistance and Training," IEO Background Paper No. BP/15/03 (Washington: International Monetary Fund).
- Ter-Minassian, Teresa, 2022, "The IMF and Capacity Development—Case Studies for Western Hemisphere," IEO Background Paper No. BP/22-02/05 (Washington: International Monetary Fund).