Revenue Administration and Public Financial Management Reform in Southeast Europe (SEE II) project

Mid-term external evaluation

December 2018 to April 2021

Evaluation Report

October 2022

consulting
BASE

CONTENTS

ACKNOWLEDGEMENTS

We have found this work on evaluating the SEE II program extremely interesting and rewarding. We have been very impressed by the very helpful and highly professional approach taken by the IMF team in this evaluation exercise.

We thank in particular Priyanka Chaturvedi and her colleagues in ICD for their support, guidance and help with logistics throughout the assignment. The coordinators for the two workstreams, Andja Komso for RA, and Arturo Navarro for PFM, provided valuable technical insights as well as help with progressing the evaluation with IMF staff and counterparts. The IMF and donor comments on the various drafts of the inception and final reports greatly assisted us in producing a more enriched final product. Finally, we would like to thank all IMF staff and experts, EC and SECO staff at head quarters and SEE II countries, and country counterparts for participating in interviews and surveys, and providing candid and insightful responses.

ABBREVIATIONS AND ACRONYMS

ALB	Albania
ALM	Asset and liability management
BiH	Bosnia and Herzegovina
CD	Capacity development support
CEF	IMF Common Evaluation Framework
CIPS	Compliance improvement plan strategy
CRM	Compliance risk management
DP	Development partners
EC/EU	European Commission/European Union
ERP	Economic Reform Programme
Evaluation	This mid-term external evaluation of SEE II
FRM	Fiscal risk management
FY	IMF fiscal year (May to April)
HQ	IMF headquarters
HWI	High wealth individual
ICT/IT	Information communication technology/Information technology
IFMIS	Integrated Financial Management Information System
KEQ	Key evaluation questions for the evaluation
KOS	Kosovo
KPI	Key performance indicator
LTO	Large taxpayers office
LTX	Regional Advisors
MKD	North Macedonia
MNE	Montenegro
MoF	Ministry of Finance
MTBF/MTEF	Medium-term Budgeting/Expenditure Framework
OBJ	IMF RBM logframe Objective
OECD	Organization for Economic Cooperation and Development
OECD-DAC	OECD Development Assistance Committee
Partners	EC and SECO
PEA	Political economy analysis
PEFA	Public Expenditure and Financial Accountability
PFM	Public financial management
PFRAM	PPP Fiscal Risk Assessment Model
PIM/A	Public Investment Management/Assessment
PPP	Public-private partnership
Program	SEE II

RA	Revenue Administration
RBM	Results-based management
SECO	Swiss State Secretariat for Economic Affairs
SEE II, or SEE Phase II	Revenue Administration and Public Financial Management Reform in Southeast Europe project
SER	Serbia
SOE	State-owned enterprise
STX	Short-term expert
TA	Technical assistance
TADAT	Tax Administration Diagnostic Assessment Tool
TOR	Terms of reference for the Evaluation

EXECUTIVE SUMMARY

The Revenue Administration and Public Financial Management Reform in Southeast Europe project (SEE II, or the Program), implemented by IMF, and funded by the European Commission (EC) and the Swiss State Secretariat for Economic Affairs (SECO), supports Revenue Administration (RA) and Public Financial Management (PFM) reform in six countries[1] in the region. The current, second phase of the Program started in December 2018, and will be completed in December 2022. The Program has a budget of approximately US$18m.[2] SEE II is intended to contribute to the EU accession process, the Addis Tax Initiative objective of improving the efficiency and effectiveness of revenue administration systems, and help align beneficiary country RA and PFM systems with international good practices. It is carefully integrated with the IMF surveillance process.

This report sets out the findings and suggestions from a mid-term evaluation of SEE II (the Evaluation), undertaken to assess the Program's overall performance in achieving its mandate, and to formulate recommendations for improvement. The Evaluation has been conducted through assessment, by applying OECD-DAC evaluation criteria[3], of a sample of capacity development (CD) projects undertaken by SEE II. The samples have been grouped into three country and two topical case studies for each workstream. Assessment has been undertaken through document reviews, interviews and surveys. The IMF proposed scoring system[4] has been applied to determine aggregated scores for assessed topics, countries, workstreams and the Program. The review covers SEE II activities between December 2018 and April 2021.

SEE II has performed very well, especially in the context of COVID-19, which, despite a quick and flexible IMF response, had a considerable impact on both CD delivery, and counterpart absorption and implementation. Overall, the Evaluation considered both workstreams to be well designed; providing very high-quality technical advice; showing strategic flexibility in response to the pandemic; making good progress (in light of COVID-19 and other counterpart constraints) in supporting reform implementation in critical areas; and building effectively on the foundations laid during the previous phase. The overall rating for the Program, and both RA and PFM workstreams, is 'good'. Both workstreams were also rated similarly for each of the OECD-DAC evaluation criteria. The evaluation ratings for the Program, which are based on a selected sample of countries and topics, are also influenced by the choice of countries, which included one where there has been limited CD implementation.

SEE II relevance was the highest rated evaluation criterion, an 'excellent'. This reflected the thorough analysis and diagnostic to identify critical gaps, and counterpart consultations, on which the workstreams based their priority areas for intervention. There has been significant buy-in from CD beneficiaries, which was evident from interviews and survey responses. The flexible response to COVID by shifting areas of emphasis, and introducing new ones, to meet emerging priorities was also a contributor to the high rating. More informed consideration of political economy and counterpart capacity constraints in intervention design could have been considered.

The Program's coherence was also assessed to be high, with a rating of 'good'. There was strong support delivered for the EU accession agenda in both workstreams, with PFM being more directly influenced by this issue. There was also apparent complementarity with IMF surveillance and other activities; counterparts' strategic agendas and institutional strategies; and close cooperation with development partners (DPs). In some areas, closer attention could have been paid in coordinating interventions with regard to their sequencing and interlinkages.

The effectiveness of both workstreams was adversely affected by the impact of COVID-19 and other constraints in counterpart countries and institutions. In addition to impacting CD delivery and

[1] Albania (ALB), Bosnia and Herzegovina (BiH), Kosovo (UVK/KOS), North Macedonia (MKD), Montenegro (MNE), and Serbia (SER).

[2] As at 31 April 2021, the respective commitments in US Dollar equivalents were EC: US$11.3m, and SECO: US$6.7m

[3] Organization for Economic Cooperation and Development, Development Assistance Committee evaluation criteria being relevance, coherence, effectiveness, efficiency, impact and sustainability.

[4] Based on responses to key evaluation questions for each evaluation criterion in the evaluation terms of reference, scores and qualitative gradings are awarded as: 4 = excellent/fully achieved; 3 = good/largely achieved; 2 = modest/partially achieved; and 1 = modest/not achieved.

monitoring in general, and to countries that couldn't easily switch to online modes in particular, COVID stretched counterpart resources as a result of restrictions as well as diverted attention (including in some cases staff reassignments) to addressing resultant emerging priorities. Counterpart institutions have also been affected by lack of resources; high staff turnover; in some cases frequent changes in senior officials; and slow decision-making. In politically sensitive areas, such as PIM/PPP, progress has also been slow, although in PFM the authorities took the opportunity to, for example, revise PFM strategies. Effectiveness was rated as in between ''largely achieved' and ''partially achieved'.

SEE II efficiency was similarly influenced by the adverse impact of COVID-19, in both CD delivery and absorption. Many of the other issues affecting effectiveness also impaired efficiency. IMF did switch to online CD and workshop delivery quickly and the throughput of CD was not significantly affected after the initial stages of COVID. However, feedback indicates that counterparts do not view online cooperation as being as productive as in-person missions and workshops. Program monitoring was also affected, with LTXs not being able to make country visits, which caused some additional loss in implementation momentum. Efficiency is rated as modest.

The Program's impact and sustainability are rated at between 'good' and 'modest' at this stage. Most of the interventions in both workstreams are long-term in nature, and it will take time for their higher-level contributions to impact to be felt. Nevertheless, building on reforms started in the previous phase, SEE II has continued institutional reforms, and added new ones, that are likely to contribute to impact and sustainability. A good indication of this is that in spite of the challenging context (i.e., Covid pandemic), human and institutional capacity was mostly sustained in beneficiary countries albeit to varying degrees. The 'pull' of EU accession is another motivator for continuing reforms, albeit the proximity to accession can influence the urgency of reforms. The relatively low effectiveness of the current phase; staffing and resource constraints at counterpart institutions; and political uncertainty in some countries pose risk to impact and sustainability.

The main recommendations from the Evaluation are aimed at the Program developing a more systematic approach to translating technical advice to outcomes. These are:

1. Undertaking formal analyses of political, governance, corruption and institutional systems and structures in higher risk SEE II countries, at both Program and project levels, to identify and address potential barriers to effective implementation of CD. In some political economy sensitive cases, such as PIM/PPP reform, CD delivery approach may need to be adjusted to incorporate more politically-oriented engagement (either directly, or through development partners) with a wider range of stakeholders.
2. Conducting resource analyses of counterpart staffing levels, skills, motivation, and softer qualitative issues such as incentives, attitudes, etc., to identify and address gaps, that could be addressed through CD, or by counterparts.
3. Shifting strategic focus of SEE III to consolidation and capacity building. In some cases this may require partnering with other development partners who have the resources to provide embedded capacity building support within beneficiary institutions (for example, by placing long-term resident advisors).
4. IMF, the EC and SECO might wish to consider a closer, more strategic partnership for SEE III, involving for example, joint dialogue with authorities, political escalation where necessary, and complementary CD delivery.
5. Project milestones and outcomes could be more usefully designed for monitoring implementation of advice, if they are more based on historic experience and counterpart capacity, constraints, and commitments to other large initiatives.

Other recommendations relate to potentially strengthening program impact, through:

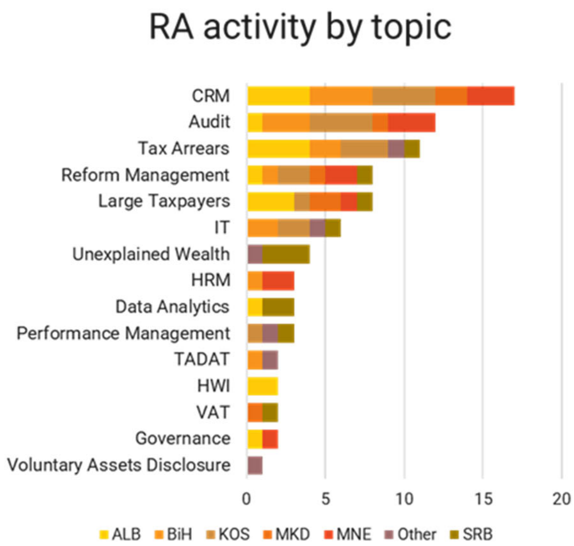
6. In keeping with the 'big picture' perspective of encouraging voluntary compliance, opportunities for optimising CRM principles across key functional areas (i.e., awareness, education and outreach) should be explored.
7. A strategic approach to data analytics should be considered that allows for data sharing and full utilisation across operational areas, for example CRM, Tax Arrears and Audit.
8. Consider engaging experts in cross-cutting themes such as anti-corruption, climate change, and gender, to review relevant program and project level design and introduce features that might enhance effectiveness in these areas.

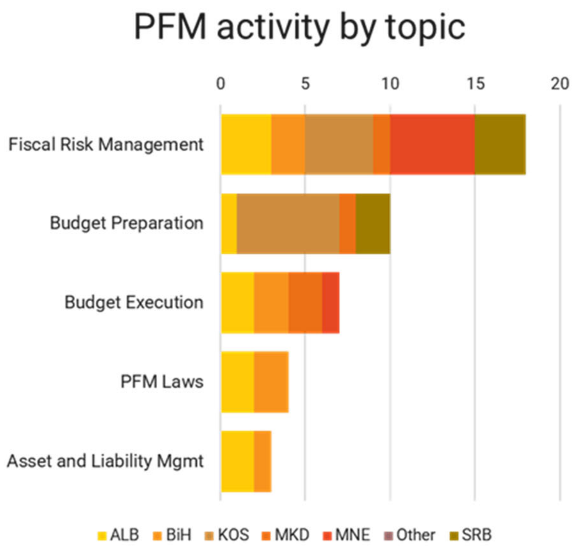
1. INTRODUCTION

1.1. PROGRAM OVERVIEW

The Revenue Administration and Public Financial Management Reform in Southeast Europe project (SEE II, or the Program), implemented by IMF, supports Revenue Administration (RA) and Public Financial Management (PFM) reform in six countries[5] in the Western Balkans. Starting in December 2018, it had a three-year budget of US$15.3m (since increased to US$18.4m), and its duration has been extended by one year to December 2022. The Program is funded by two partners, the European Commission (EC) and the Swiss State Secretariat for Economic Affairs (SECO) (together, Partners). SEE II is intended to contribute to the EU accession process, and the Addis Tax Initiative objective of improving the efficiency and effectiveness of revenue administration systems, and help align beneficiary country RA and PFM systems with international good practices. The SEE II expected outcome (in conjunction with other RA and PFM reform programs) is that all beneficiaries will have improved the institutional capacity to meet the EU accession criteria. A steering committee, comprising representatives from IMF and the Partners oversees the Program, providing strategic guidance and assisting in setting priorities for program activities.

SEE II activities have built on earlier work and are based on findings of systematic diagnostics. The previous phase of the program, funded by EC (PFM) and SECO (RA), laid the foundation for the current work. In addition, various diagnostics, such as Tax Administration Diagnostic Assessment Tool (TADAT), Public Expenditure and Financial Accountability (PEFA) assessment, and Public Investment Management Assessment (PIMA), combined with demand from capacity development support (CD) beneficiaries, have helped defined the scope of SEE II. Although the remit for each workstream is wide, in practice certain areas, such as fiscal risk management and budget preparation for PFM, and compliance risk management, audit and arrears for RA, account for a large proportion of SEE II activities[6]. Activities of the Program are also informed by the IMF surveillance process, and are closely integrated with it. For example, during the review period one of the SEE II countries benefited from an IMF Policy Coordination Instrument program, with IMF-supported tax administration and PFM reforms being an essential part of the program





Source: IMF

[5] Albania (ALB), Bosnia and Herzegovina (BiH), Kosovo (UVK/KOS), North Macedonia (MKD), Montenegro (MNE), and Serbia (SER).

[6] Activities are broadly defined as missions, or a set of missions for with a technical assistance report is produced.

SEE II CD is delivered to counterpart institutions through a variety of channels. These include IMF headquarters (HQ) staff; regional advisors (LTX); and short-term experts (STX). Thematic regional workshops are also held. Since early 2020 until the end of the evaluation period, SEE II missions have been undertaken virtually, and the workshops held as webinars. As at the end of fiscal year (FY) 2021, 51% of the total SEE II operating budget of US$17.2m[7] had been spent, US$5.4m (62%) on RA activities, and US$3.3m (38%) on PFM. These correspond to 84 CD activities or missions for RA, and 42 for PFM.

1.2. SUMMARY OF EVALUATION APPROACH AND METHODOLOGY

The objective of the mid-term evaluation of SEE II (the Evaluation) is to assess the overall program performance in achieving its mandate as envisaged in its Program Document, and to formulate recommendations for improvement. The terms of reference for the Evaluation (TOR) set out the evaluation scope, methodology, work program and deliverables. In particular, it provides that the review is to be conducted with reference to the six OECD-DAC evaluation criteria of relevance, coherence, effectiveness, efficiency, impact and sustainability. A set of 30 key evaluation questions (KEQ), categorised among the evaluation criteria, are provided, to guide assessment of projects[8]. An approach to sampling projects is proposed, along with a methodology for rating and scoring them for each of the evaluation criterion. The TOR provide three further questions to guide the assessment of strategy and operations management of SEE II[9].

This report presents the findings of the Evaluation, covering the period since the start of the current phase, in December 2018, to April 2021. Many of the findings from interviews and surveys reflect the current state of implementation of CD delivered during the evaluated period. The assessment presented in this report is based on a review of 25 projects (10 for RA, and 15 for PFM), undertaken in clusters as case studies to draw wider lessons. Three case study countries were selected, the same ones for PFM and RA, and two topical areas for each of the workstreams: PIM and public-private partnerships (PPP) for PFM; and compliance risk management (CRM) and arrears management for RA.

A range of evaluation tools have been deployed to undertake project assessments. Extensive documents have been researched, including those related to SEE II governance, management and reporting; project design and CD advice and associated results-based management (RBM) logframe outcomes, milestones and indicators; strategies, policies and related reports produced by beneficiaries; relevant IMF policies and other surveillance and other reports; information related to the previous SEE phase; and background information about the case study countries. Interviews have been undertaken with IMF staff; LTX and STX; EC and SECO head quarter staff, and program managers from case study countries; and staff from counterpart institutions. A total of 77 individuals were interviewed, some more than once. The Evaluation also relied on surveys of CD beneficiaries; SEE II workshop participants; IMF staff; and development partners (DP), including SECO and EC local delegations. A total of 79 responses were received.

The IMF approach to rating projects for each of the OECD-DAC criteria has been applied. Each criterion has been awarded a score and a qualitative grading between 4 and 1 based on an assessment of the responses to the relevant KEQ. These are, 4: 'excellent', or 'fully achieved'; 3: 'good', or 'largely achieved'; 2: 'modest', or 'partially achieved'; and 1: 'poor', or 'not achieved'. Scores have been averaged on an unweighted basis to arrive at aggregates for each project, country study, topical area, as well as (as presented in this report), all countries, all topics, all objectives, the RA and PFM workstreams, and the Program overall.

The remainder of this report is divided into five sections: Section 2 presents an overview of findings for the Program; Sections 3 and 4 present the evaluation findings for the RA and PFM workstreams respectively; Section 5 provides our assessment in response to the evaluation questions on SEE II strategy and operations management; Section 6 reviews SEE II workshops and webinars; and Section 7 concludes the report with a presentation of the Evaluation's main findings and conclusions.

[7] Pre IMF Trust Fund fee

[8] Projects are defined as an objective, as set out in the SEE II results-based management logframe, in a country.

[9] TORs provide that SEE II strategy and operations management assessment is not to be rated.

2. ASSESSMENT RESULTS AT SEE II PROGRAM LEVEL

OECD-DAC aggregate scores[10] for workstreams and Program			
	RA	PFM	Program-level average
Relevance	3.8	3.4	**3.6**
Coherence	2.9	3.2	**3.1**
Effectiveness	2.5	2.4	**2.5**
Efficiency	2.3	2.3	**2.3**
Impact	2.2	2.3	**2.3**
Sustainability	2.2	2.4	**2.3**
Overall	**2.7**	**2.7**	**2.7**

The aggregate score for the average of OECD-DAC ratings at the Program level is 'good'[11]. Both the RA and PFM workstreams have the same average score, although there are variations in scores attributed to individual OECD-DAC evaluation criteria between them. The overall score reflects a strong performance of the Program, despite the adverse effects COVID-19, which affected the majority of the period under review.

A number of common issues had important impact on the performance of both SEE II workstreams, greatly influencing their assessments. The pandemic affected the supply of CD with travel restrictions, with the move to CD delivery online generally viewed by beneficiaries to be less effective than in-person interactions, both with experts and also during workshops. It also caused disruptions at beneficiary institutions, with the RA workstream being affected more given its operational focus. Other factors that influenced performance, both positively and negatively, included: i) both workstreams built their work programs on strong foundations laid in the previous phase; ii) the work programs were designed robustly, guided by detailed diagnostics such as TADAT (for RA) and PIMA (for PFM); iii) very high technical quality of advice was delivered by IMF, engendering confidence in counterparts and willingness to translate CD support to action, constraints permitting; iv) similar CD delivery modalities were used, with a combination of IMF HQ missions, LTX, STX and workshops complementing each other; v) SEE II countries are motivated to act by the prospects of accession to the EU, although the strength of influence of this issue was at least partly affected by the advancement of negotiations, with PFM more directly affected by this factor; vi) both workstreams reacted flexibly to the consequences of COVID-19, either by changing emphasis between topical areas, or introducing new, relevant activities; vii) the political economy of the countries, some of which saw a number of changes in government, or other political disruptions, impacted a number of projects, with the PFM workstream being more affected (although resulting changes in senior counterparts also affected RA); viii) resource and capacity constraints affected both RA and PFM, although given the stage of implementation of the work programs, RA was impacted marginally more; ix) both workstreams coordinated well and leveraged the activities of DPs, which had a particularly strong positive impact on PFM performance, given the wide scope of its remit and greater opportunities for such cooperation; and x) assessment ratings have been affected by the choice of sample countries in particular, with one country, selected for lessons on low implementation experience, scoring significantly lower than the others in both workstreams.

Although the average OECD-DAC score assessed for both RA and PFM workstreams are the same, variations in ratings of the different evaluation criteria reflect features that differed between them at the margins. The next two sections of this report detail the rationale behind the assessment ratings for each workstream. A wide range of factors positively and negatively have affected the ratings for the evaluation criteria for objectives, case study countries and topics, that are distilled into the small net differences in scores highlighted in this section. Nevertheless, some the most common influences on workstream ratings and differences between them have been: i) political economy issues depressed the relevance rating of PFM more than RA; ii) PFM's coherence rating was conversely more positively influenced by donor coordination issues; iii) the impact of COVID-19 had a large dampening influence on the performance, and average assessment ratings, of the effectiveness and efficiency evaluation criteria, although there were important variations in scores within the workstreams between countries and RBM

[10] For each project assessed, an average of the OECD-DAC criteria scores has been calculated to provide an overall project-level score (based on the methodology in the footnote below). The scores for all projects within each country, topic and objective (for each OECD-DAC criterion and overall project scores) were averaged to provide a set of individual country, topic and objective level scores. This exercise was repeated for all countries, topics and objectives to arrive at the overall country and topic scores for the assessed portfolio. Finally, the criterion score for all projects were averaged to arrive at the assessed portfolio-level scorings and ratings for each workstream. Simple averaging has been applied for aggregating all scores. Average scores have been rounded up or down to the nearest integer to attribute a qualitative rating.

[11] Evaluation ratings are, 4 = 'excellent'/'fully achieved'; 3 = 'good'/largely achieved'; 2 = 'modest'/partially achieved'; and 1 = 'poor'/'not achieved'). Target-related criteria, relating to degree of achievement, is used for rating effectiveness. More qualitative ones, excellent, good, etc., are used to rate the other OECD-DAC criteria.

objectives. The effectiveness rating of the RA workstream is likely to have been marginally more affected by COVID; iv) the impact rating is strongly influenced by the score for effectiveness, and sustainability additionally by similar organisational issues in the sampled countries, although PFM has scored marginally higher for these evaluation criteria given more direct linkage between the workstream topics and accession requirements, and the additional impetus that might create to implement and sustain reforms.

The evaluation ratings for SEE II are comparable to another recent evaluation, that of the Managing Natural Resource Wealth Thematic Fund (MNRW TF), that covered the COVID-19 period. The overall score for MNRW TF was also 2.7, and the average rating for its PFM and RA workstreams was also 2.7. The average scores for the two workstreams for the OECD-DAC criteria were: relevance, 3.5; coherence, 3.3; effectiveness, 2.4; efficiency, 2.6; impact, 2.2; and sustainability, 2.2.

3. ASSESSMENT OF REVENUE ADMINISTRATION WORKSTREAM

3.1. Key achievements

The overall assessment of the Evaluation is that the IMF's RA portfolio has performed reasonably well against its outcomes and objectives. CD support under the RA workstream has built on earlier reforms in strengthening organisational arrangements and capacity across the RA agencies. Similarly, SEE II CD support has contributed towards the implementation of Economic Reform Programme (ERP) reforms required to support the EU accession process and is aligned to key reform priorities identified through IMF surveillance of beneficiary countries. Significant achievements to date include:

- The development of improved processes for modernisation planning and consolidation of organisational arrangements. These reforms have provided beneficiary counterparts with fundamental structures required to carry out effective tax administration functions.
- The consolidation of debt collection functions and approaches in line with international good practice to establish a credible debt management program, that facilitates an efficient collection system within beneficiary RAs.
- The implementation of modern CRM approaches and methodologies to tackle systemic compliance risks within beneficiary administrations.
- The improvement of audit capabilities and procedures through the implementation of various audit techniques and strategies, to support the undertaking of effective and efficient tax audits.

There have been measurable improvements in the RAs of the SEE II countries which has been captured under various assessments. With the exception of 2020, average tax-to-GDP ratios increased across all six SEE II beneficiary countries in the period under review. TADAT assessments have also been very useful in identifying weaknesses in core tax administration areas, including those in which enhancing capacity was required. In the case of SEE II countries, persistent weaknesses and capacity deficits were identified under Performance Indicators relating to topical areas such as CRM, Arrears, Refunds, Audit and Data. The project design has drawn heavily on the findings of TADAT assessments and has supported the deployment of targeted interventions across the various administrations to address these weaknesses. Nonetheless, the lack of adequate data on recent TADAT diagnostics has meant that a comparative assessment against which progress between intervening assessment periods could not be tracked.

3.2. RA workstream-level assessment

RA workstream	
Overall	
2.7	
Country	Topic
2.6	2.7
OBJ 1	OBJ 2
2.8	2.5

The overall assessment of this review is that the IMF's RA portfolio has performed reasonably well against its objectives. The two RA RBM logframe Objectives (OBJ) targeted under SEE II were evaluated through a sample of projects in three country case studies and two topical areas, CRM and Arrears, assessed in five SEE II countries. Aside from the selected cross cutting topics, this evaluation covered as part of country case studies projects related to Large Taxpayers, High Wealth Individuals, IT, Performance Management/Key Performance Indicators (KPI), Reform Management, and Audit. The average score across all OECD-DAC evaluation criteria (comprising all three case study countries and the topical areas) for each Objective were: ***Objective 1*** *- Better Revenue Administration, management and governance arrangements,* 2.8[12] ('good'); and ***Objective 2*** *- Strengthened core tax administration functions*, 2.5 (between 'good and 'modest').

[12] Evaluation ratings are, 4 = 'excellent'/'fully achieved'; 3 = 'good'/largely achieved'; 2 = 'modest'/partially achieved'; and 1 = 'poor'/'not achieved').

KEY FINDINGS FROM THE EVALUATION OF SEE II RA WORKSTREAM
What has been done well and has been successful
▪ **The evaluation finds strong evidence that the IMF CD RA interventions are based on diagnostic assessments aimed at strengthening capacities and addressing core weaknesses in tax administration.** The IMF's CD approach to RA emphasizes the role of evidence-based diagnostic assessments such as TADAT and information from surveillance activities as foundations for the design and implementation of effective project interventions. The particularly strong portfolio-level results against the OECD DAC criteria for Relevance and Coherence supports this assertion. Further the clear alignment between the IMF CD interventions and the beneficiary country strategic/operational reform plans implies good stakeholder engagement in the diagnostics process. ▪ **Overall, the technical quality of CD interventions is of high standard reflecting the normative role played by the IMF with respect to RA reforms globally**. Important reform aspects such as alignment to EU taxation norms, other best practices, as well as modern concepts and approaches particularly in the treatment of tax arrears and addressing of CRM issues have been incorporated into CD interventions. ▪ **Broadly, there has been appropriate sequencing of RA reforms**. The evaluation finds strong evidence of the IMF's use of the TADAT tool to promote in-country consultation and dialogue with stakeholders in the beneficiary countries. This has supported the appropriate sequencing of CD interventions including the sequencing of various CD delivery techniques. During the Covid-19 pandemic, CD delivery modalities and activities were adapted to match ad-hoc needs that arose amongst beneficiary countries. ▪ **The pilots undertaken during CD interventions were highly beneficial and increased the efficiency of capacity development measures**. Pilot programs provided RA staff with hands on experience ensuring that staff acquired requisite skills and were able to test and 'try out' these skills in counterpart demand-driven sectors. ▪ **The use of regional events and peer to peer learning networks has had a positive impact** allowing RAs to engage in knowledge sharing with each other and to benefit from similar reform experience. ▪ **There is also sufficient evidence that the extent of adoption of reform recommendations made under the project were tracked** across the various countries.
Areas for potential improvement
▪ **Although the project-level objectives and outcomes framework follow a consistent structure, the wording of certain milestones appear vague and ill-defined;** making it difficult for these milestones to be effectively monitored or measured against program objectives and in some cases being beyond the control of the Program**.** Careful crafting of the wording for milestones should be done to ensure they consistently meet the SMART criteria. As well, the timelines for the achievement of some milestones appear to be over optimistic and challenging to achieve. ▪ **While there is sufficient evidence that efforts are made to support dialogue and collaboration with beneficiary country stakeholders particularly at the start of the CD program. I**n some instances, stakeholder buy-in and understanding of some reforms pursued were inconsistent and insufficient**.** Greater efforts should be made to create a common understanding and buy-in across multiple stakeholders through the incorporation of learning or reflection points throughout the program. This involves reflecting on what's working, what's not, what's changing within the country context including what windows of opportunity continue to exist to pursue planned reforms. ▪ **Even though CD delivery in most of the beneficiary countries exhibited a number of approaches (e.g., close involvement of key stakeholders in the TADAT diagnostic process; a focus on capacity-building; coordination with other DPs operating the RA reform space etc.),** that would normally constitute a change management strategy; an explicit commitment to or adoption of a change management strategy has been lacking. Given the complexity of some of the RA reforms pursued; the level of sustained beneficiary country commitment required to push through certain reforms; and the potential constraints that usually require addressing; a more formal approach to change management would be more effective in managing such challenges. Within the context of program design, this would mean ensuring that appropriate technical interventions exist to build support for 'change' as well as to enhance capacity for managing 'change' within beneficiary RAs. ▪ **While the quality of reporting is generally of a good standard and commendably concise; more consistency is required in the format and content of reports.** Equally, for a more comprehensive picture of CD interventions undertaken, reports should follow a more logical structure that includes information with respect to overall progress made, reasons for lack of progress, and the remedial changes or variations made as a consequence. ▪ **There is room for argument that CD interventions in some countries may have been over ambitious in scope**; although certain aspects of the program operated in areas of revenue reform (e.g., in the area of CRM) which are typically slow to deliver strong results**.** Whilst some countries demonstrated substantial reform progress, in others the reform pace was slow for a number of reasons. They include political instability, economic deficiencies of country systems, inadequate organizational structures coupled with insufficient technical or management skill levels within some RA agencies. Such characteristics constitute significant risks to reform impact that should be complemented with adequate mitigation strategies. It is worth noting that although the various program documents provide a high-level risk matrix for each country, the lack of granular detail makes it difficult to evaluate the quality of the ongoing risk assessment process.

For the most part, Objective 2 scores are lower in most of the criteria assessed and this can be attributed to a number of factors. The lower overall rating under Objective 2 can be explained in part by spill over effects from the outbreak of COVID-19 on certain areas of tax administration. Aside from the pandemic, features in the political economy in some countries, including transfers or reallocation of beneficiary institution staff; a refocus of reform priorities in some; and reasons outside of the IMF control such as law changes in certain countries, influenced the effectiveness of organizational structures, all contributed the lower scores attributed to Objective 2 and its underlying outcomes.

Country performance across the RA portfolio showed some variation in achievement, with the overall rating being close to 'good'. Whilst relevance and coherence remained high, the unanticipated advent of the pandemic adversely influenced effectiveness, impact and sustainability to different degrees in all three case study countries. This resulted in project interventions being delayed and, in some instances, some milestones required readjustments to their achievement timeline. Some of the delays were mitigated by remote CD delivery, however this incentive was only taken advantage of in certain countries where appropriate structures and adequate level of digitalization existed to support remote working. The average of OECD-DAC criteria scores from the country studies was 2.8. The three case study countries scored 2.9, 2.8, and 2.5, respectively, with the latter country scoring lower than the other two under all evaluation criteria except for relevance and coherence.

The overall performance of the Program in the topical areas of Arrears and CRM is more moderate. Performance is broadly consistent across the six evaluation criteria with no significant variation in scores across the various countries. Arrears had an overall score of 2.5 and CRM had an overall score of 2.8 (to different degrees between 'good and 'modest, with Arrears in particular showing more limited progress in implementation and outcomes in most countries. A number of factors responsible for this and which vary from country to country include: poor IT supporting systems, slow rate of progress major IT implementation projects, inadequate legislative frameworks to address core operational areas such as uncollectible debt, along with a growth in tax arrears as result of the Covid pandemic, Aside from the impact of the pandemic, it's worth highlighting that reforms in these areas must necessarily adopt a longer-term view. Sustained success in these topics usually requires considerable institutional change and embedding of practices, which realistically only happens gradually. Typically, they also require integration with wider policy and operational reforms which can serve as reform enablers. There is also the fact that most of the SEE II countries were either transitioning to or implementing new Information Communication Technology (ICT) systems, which in some instances contributed to delaying reforms and limited the bandwidth available in certain administrations to pursue other reforms.

3.3. Relevance

Relevance	
Overall	
3.8	
Country	Topic
4.0	3.7
OBJ 1	OBJ 2
3.9	3.8

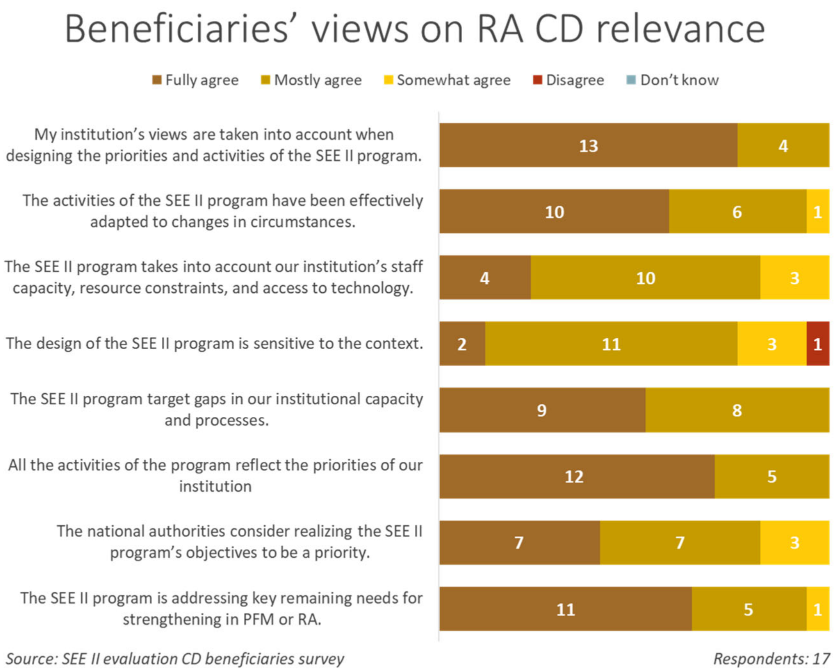
The RA workstream has performed particularly well against the relevance criteria, in that it addressed critical capacity gaps, was largely tailored to the stated policies and priorities of beneficiary governments, and was sufficiently adaptive to environmental changes, particularly those related to COVID-19. The overall workstream level score for relevance is 3.8, with a rating of 'excellent', and this was consistently achieved in respect of both Objectives. The overall relevance score was marginally affected by some adverse ratings for both topical areas in two different countries.

The evidence-based approach to project design and consultation with stakeholders is very strongly present in the assessed projects. Information from TADAT diagnostics and other IMF surveillance activities provided robust and objective insights that were explicitly used to define the focal areas for CD interventions. The Program is building capacity in core tax administration functions and has supported beneficiary institutions develop modern organizational structures and sound processes, in line with international good practice. Further work is needed to consolidate these reforms, and the CD approach should balance reform ambition with the capacity of beneficiary countries to ensure meaningful and sustainable RA reforms in priority areas.

The RA portfolio showed a very high level of relevance in relation to its alignment with the stated policies and priorities of the beneficiary governments. In all countries reviewed the project objectives are directly aligned with specific goals stated in the relevant RA reform strategies and plans. The Program has also considered the political and economic context to an extent. Nevertheless, a more systematic methodology could be applied to understand the political economy of reforms, both at the contextual level and, in particular, the project level. Periodic real-time political economy analysis and risk monitoring would benefit the direction and management of interventions, enabling proactive action. One of the case study countries scored lower than others across all implementation-related criteria, and political-economy issues appear to have been a key barrier in achieving results.

The Program was appropriately flexible to adapt to the COVID crisis, both in respect of adapted milestone achievement timelines, as well as *ad-hoc* support in targeted areas as a direct response to beneficiary requests. CD was quickly shifted to remote/online delivery, although uptake varied between countries, depending on the level of adaptability and digitalization within each RA.

Relevance was high across both topical areas and across all countries. CD interventions delivered addressed critical gaps in Arrears and CRM identified under diagnostic assessments such as TADAT; and were generally reflective of the stated reform priorities of RAs and ongoing tax administration reforms. In both topical areas, CD interventions addressed weaknesses in operational capacity, organisational structures, and operational application of good practice in Arrears and CRM through the development of solutions, methodologies and procedures.



RA CD beneficiaries' survey responses confirm the high relevance of the Program. A significant proportion feel their views are taken into account; that CD is addressing key priorities; and that delivery has been responsive. However, a fewer proportion fully agreed that the Program takes into account implementation constraints; that it is sensitive to the context; or that national authorities share in the institution's prioritisation of SEE II activities.

3.4. Coherence

The RA portfolio has also performed well against the coherence criteria, in that there was reasonable compatibility between CD objectives and those of beneficiary institutions; CD interventions were consistent with IMF surveillance and diagnostic activities; and collaboration and coordination with other DPs operating in the reform space was done. The criterion achieves a score of 2.9 (rating of 'good') across all reviewed categories.

Coherence	
Overall	
2.9	
Country	Topic
3.0	2.9
OBJ 1	OBJ 2
3.0	2.9

Internal coherence was largely achieved in that CD objectives were mostly in alignment with the reform objectives of the various RAs, as well as with strategic reform commitments of various beneficiary governments. This happened particularly with respect to actions under the ERPs agreed with the EU in support of the accession process and those recommended under IMF surveillance activities, such as Article IV consultations. Coherence of project design was given particularly close attention during program design, but less so during implementation, partly reflecting (in some cases) the changes to reform priorities mid-implementation on the beneficiary side.

The alignment of the IMF's CD interventions with other DPs such as the EU and the World Bank is commendable and complementary. Particular attention is paid at program design stage to record and take into account the activities of other DPs and routine DP coordination activities are carried out in-country through various meetings. A clearer reference of the coherence and synergies of the RA portfolio with other reform interventions implemented by other DPs at country level would be useful.

Coherence was relatively achieved in both topical areas across all countries. CD interventions were well coordinated at the country level, resulting in a good level of counterpart buy-in. Coherence was however challenged on a number of fronts. In the topical area Arrears, coherence was affected by the lack or slow rate of progress in ongoing complementary reforms (e.g. IT implementation reforms) in a number of countries. In the topical area CRM, the integration of certain interventions while present was not wholly consistent (e.g. linkages between CD interventions targeted at enhancing compliance in Arrears and Audit).

The RA CD beneficiaries survey confirms coherence-related findings from other sources. 71% strongly agreed that CD recommendations are consistent with other IMF advice; 65% that they are consistent with, and help the country's broader economic program and EU accession agenda; while 53% that CD is effectively coordinated with other DPs. The balance all agreed in respect of the first two questions, while for the third question, 35% agreed, and 12% somewhat agreed.

3.5. Effectiveness

Effectiveness	
Overall	
2.5	
Country	Topic
2.5	2.5
OBJ 1	OBJ 2
2.5	2.5

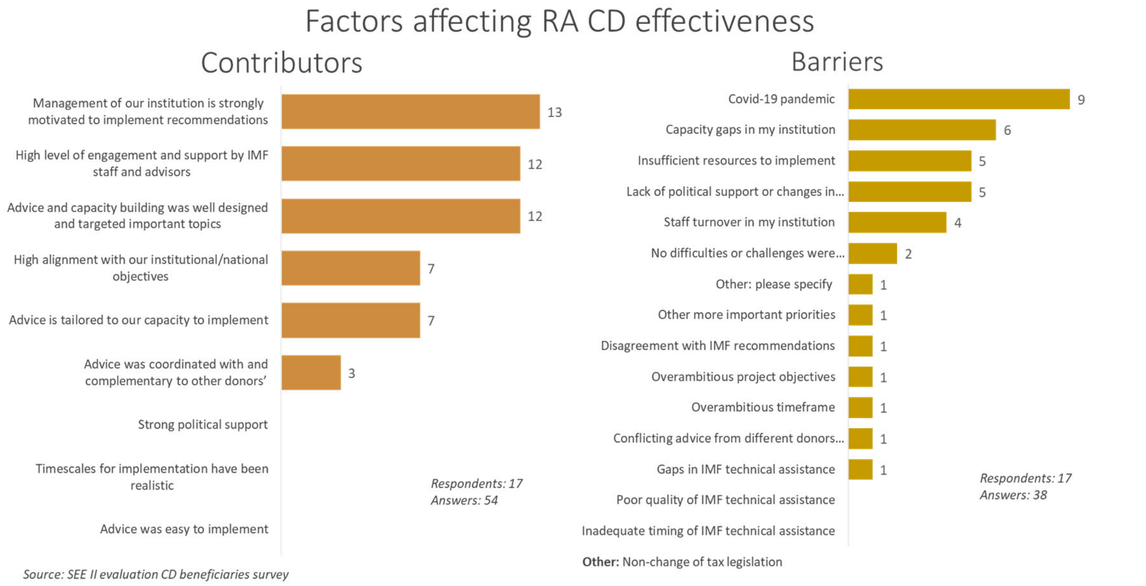
The portfolio has performed moderately against the effectiveness criterion as some results were achieved, with the majority of CD interventions having made some progress towards the achievement of objectives (to varying degrees across the beneficiary countries); although the outbreak of Covid-19 had a significant impact on the Program as a whole. The overall score for effectiveness, at 2.5, falls in ratings of between 'good' and 'modest'.

A similar level of performance was achieved against both Objectives and this is attributable to a variety of factors across the assessed countries. In particular, under Objective 1, good progress was made in the establishment of sound organisational structures and processes that build on previous reform efforts under SEE I to modernise RA structures across the beneficiary countries. At the milestone level, individual ratings across objectives demonstrate that advances have either been made or that reform steps were initiated. In other cases, this review judges it to be highly likely that RA related milestones will eventually be delivered in full, although some level of uncertainty should be accounted for in some country reform contexts. Overall, the majority of beneficiary countries and RA institutions demonstrated a good level of buy-in reflecting the sound processes used in program design and preparation. . For some countries, delays to reforms under Objective 1 affected the delivery of dependent reforms under Objective 2, resulting in less CD activity being undertaken under the latter objective than previously envisaged. Other reform constraints in this area that adversely affected the level of effectiveness included inconsistent senior management commitment to reforms in some instances, and spill over effects from the outbreak of the pandemic such as growth in tax debt and an increased inability of some taxpayers to pay taxes, ultimately limiting the extent of reforms that could be progressed under this objective.

There were varying levels of effectiveness of the Program across the beneficiary countries. COVID-19; changes in governments; in some instances changes to country reform priorities and reallocation of administration staff; and delays in the implementation of complementary reform efforts all limited the degree of effectiveness that could be achieved.

The Evaluation's scores for effectiveness are broadly in line with those self-assessed by IMF. In general, the main reason for the increase in scores of certain objectives and outcomes is due to the recognition of progress that had been made since they had been last assessed by IMF. A smaller number of objectives and outcomes received lower scores primarily due to, difficulties with verifying or validating progress with certain beneficiary RAs that had been reportedly made.

Effectiveness was moderate in the topical areas, with good progress noted through the initiation of reforms, an indication that there are good prospects for the Program to make a future impact. In Arrears, CD has been effective in supporting countries establish adequate organisational structures in line with good practice. In CRM, SEE II interventions have demonstrated strong effectiveness in the development of CRM Compliance Improvement Planned Strategy (CIPS) and sectoral risk profiling. Although the quality standard of CD interventions has been commendable, effectiveness achieved across both topical areas varied between countries, indicating the contextual factors at country level that play a significant role in shaping reforms. Effectiveness in topical areas was hampered by a number of factors. Internal factors included resource adequacy issues in some beneficiary institutions; lack of or limited data exchange, poor data storage and management; inadequate data analysis capability; and in some instances institutional resistance to reforms. External factors were related to COVID-19, which restricted operational activities in RAs, limited economic activity across countries, and reduced the ability of taxpayers to pay taxes.



Surveyed beneficiaries attribute the success in SEE II RA CD delivery almost equally to their management and IMF. Management motivation was most frequently cited contributor to CD-related successes, closely followed by high IMF engagement and design of CD. Responses to the IMF staff and LTX survey was similar, although IMF engagement and design was rated the highest equally, followed by strength of beneficiary institutions' management. Covid-19 was the most frequent cause of challenges and delays by beneficiaries, followed by institutional factors like capacity gaps, inadequacy of resources and staff turnover, as well as lack of political support. Covid-19 was rated as a much lower contributor to challenges in the IMF survey, ranking sixth. Capacity gaps and insufficient resources were ranked the two most important factors, followed by overambitious timeframes. The DPs survey confirmed the difficulties in achieving results in the region, with none saying it was very easy; 20% reasonably easy; and 80% quite difficult (none said it was very difficult). A relatively high proportion of them (40%) disagreed that Covid-19 had played a significant role in reform slow-down; the same proportion disagreed that management of RAs are strong; and a majority (53%) that the RAs are adequately resourced to implement reforms.

3.6. Efficiency

Efficiency	
Overall	
2.3	
Country	Topic
2.3	2.3
OBJ 1	OBJ 2
2.6	2.0

The portfolio is assessed to have performed moderately against the efficiency criteria as CD technical quality was high and delivery modality was reasonably appropriate to support the achievement of intended results, with efforts made to adjust delivery to the onset of COVID-19. The overall rating for efficiency, at 2.3, falls between 'good' and 'modest', at a level slightly lower than effectiveness. In spite of the pandemic and in some instances challenging political economy and institutional settings the project had to operate in, some major reforms were fully achieved and in other areas were initiated. This is due in part to the quality of CD provided by IMF and the various channels used to do so, that increased the efficiency of capacity development measures.

The portfolio's delivery modality was varied and appropriate such that technical inputs were completed to good levels. Regional workshops were particularly useful in terms of enhancing cooperation as well as promoting knowledge exchange and learning. A regional approach in this regard can also provide the critical mass required to advance the RA reform agenda across the region. CD delivery modality was also responsive and proactive. Although remote CD delivery provided new outlines of engagement, this came with its own set of uptake challenges; along with the recognition that in some country contexts CD delivery at a distance could be problematic. Likewise mobility restrictions during the Covid pandemic reduced LTX in-country visits, and the level of monitoring that could be conducted; limiting the overall CD efficiency.

Efficiency was achieved on a moderate basis in the topical areas, although the disruption from the pandemic was a significant factor affecting efficiency. Interestingly, although the challenges experienced in pursuing reforms under Arrears and CRM were similar in nature, it is the extent to which reforms were progressed under each area that is responsible for the difference in scoring between them (CRM scored an average efficiency of 2.6, while Arrears scored 2.0). The majority of CD interventions were initiated across all countries with additional support provided in direct response to demand from beneficiary countries. Nevertheless, progress was limited at the same time, as delays were experienced in some countries as they refocused certain reform priorities. Similarly, while CD support was efficient in terms of adapting to remote online delivery to counter the Covid-related mobility restrictions, .some efficiency gains were lost since not all country RA or institution contexts proved suitable for remote CD delivery.

3.7. IMPACT AND SUSTAINABILITY

Impact	
Overall	
2.2	
Country	Topic
2.0	2.3
OBJ 1	OBJ 2
2.5	1.9

The RA portfolio has performed moderately against the impact criteria in that reforms were initiated and some progress was achieved in selected areas, mainly at the institutional level. The overall score for impact, at 2.2, is rated as 'modest'. The SEE II RA workstream has had an impact overall, and this is evident in the general macro-economic condition and projections for the majority of beneficiary countries, which are cautiously optimistic amidst an uncertain economic outlook. At workstream level, progress is more visible across several areas: from building capacity, improving processes and systems, to developing appropriate policies. Likewise, of the reports reviewed under Objective 1, there was a greater proportion of 'other topics' covered than under Objective 2. On average CRM topics covered under Objective 1 represent 60% and 'other topics' represented 40% . Under Objective 2, topics covered under Arrears represent 90% while 'other topics' represented 10%. This proportional difference meant that even if CRM were to underperform, good progress made in the 'other topics' covered was likely to reflect positively in the overall impact and sustainability of reforms under Objective 1. In a number of areas, impact of IMF CD will take time, and may be difficult to attribute, due to the nature of certain tax reforms. Similarly certain interventions had not been completed at the time of the review and have only delivered partial results and in some cases no result within the review period.

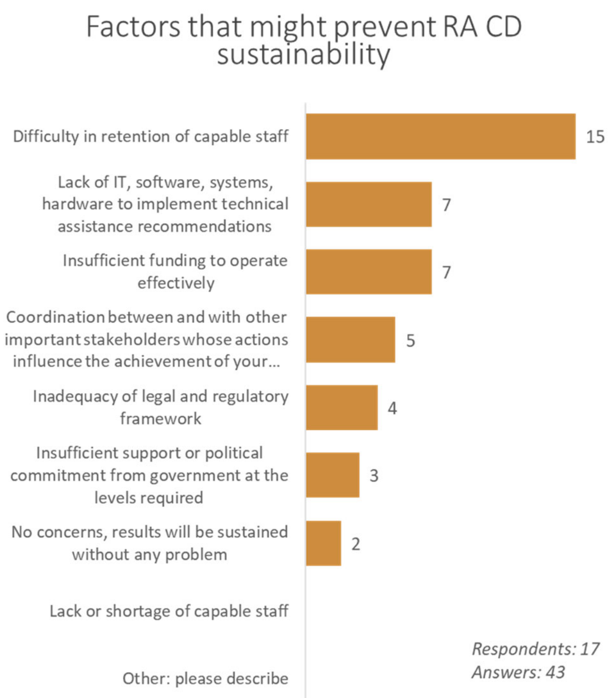
Sustainability	
Overall	
2.2	
Country	Topic
2.0	2.3
OBJ 1	OBJ 2
2.5	1.9

Impact was partially achieved in the topical areas due to the efficiency and effectiveness factors already highlighted, particularly in the context of COVID-19. It is apparent that some progress has been made overall in both topical areas, which has laid the groundwork for making future progress at the impact level. More advances have been achieved in CRM in comparison to Arrears, reflecting the unintended negative effects of the pandemic coupled with the inherent differences in the administration of both areas. In some countries staff covering Arrears were even reassigned to other tasks There were also problematic areas under Arrears such as 'uncollectible debts' which saw little to no progress achieved. Further operationalisation of reforms initiated combined with adequate performance management will serve as better indications for assessing impact of interventions in the medium-term.

The workstream has performed relatively well on sustainability, the main drivers being the introduction and implementation of RA structures and processes in line with good practice, including the provision of relevant and adequate capacity building. Overall, the performance of the workstream against sustainability is even across the beneficiary countries, with the progress achieved to date being a reasonable indicator that sustainability is likely to increase by the end of the project.

A strong emphasis was placed on capacity development during SEE II program design with the approaches to CD delivery being carefully designed for the baseline context in which CD interventions take place. However, the review identifies the extent to which the context varies across the beneficiary countries and the attendant implication of such differences. In some countries, there is a strong commitment, strategy and institutional framework to pursue RA reforms. In others, unpredictable systems of political, organisational, and individual behaviours and incentives, can impact program sustainability and requires sophisticated management and flexible systems to support testing, learning and adaptation.

Although the likelihood of sustainability has been increased by the embedding of reform practices and the development of individual capacities, adequate legal or policy backing is also required to maintain and sustain adopted reforms. There is a risk that initial benefits arising from individual-oriented topic-based training (addressing weaknesses in core tax administration areas) may be eroded over time, especially in a context of significant staff turnover. In this regard, institutional capacity development in the form of well-defined RA capacity-building strategies, based on a thorough diagnostic of capacity gaps, which provide for systematic and sustained training of staff, will be important in enhancing the sustainability of capacity development activities.



Source: SEE II evaluation CD beneficiaries survey

Sustainability was moderately achieved in both topical areas and this judgement is based on a number of indicators. The strong design of the CD interventions, the realisation of emerging results, the largely appropriate implementation modalities used to deliver CD combined with ongoing commitment and reform ownership across most countries all serve as positive indicators. Key sustainability risks such as the internalisation of reforms as well others (such as resource adequacy) that are outside the IMF's control however require addressing.

CD beneficiaries surveyed are moderately optimistic about achieving sustainability of SEE II RA reforms. Although none fully agreed to a statement that they could independently maintain or build on IMF CD, 53% mostly agreed; 35% somewhat agreed; and only 1 person (6%) disagreed. Staff retention, lack of ICT, and funding are in their opinion the most likely threats to sustainability. IMF staff and LTX were more optimistic, with 22% fully agreeing; 44% mostly agreeing; and 22% somewhat agreeing. Inadequacy of legal framework was the most frequently cited reason by them for lack of sustainability by them, followed by shortage of staff and staff retention; and motivation. The legal framework weaknesses highlighted in the IMF survey require further pursuing as part of the next phase.

3.8. OUTCOMES FROM RA CD DELIVERED IN THE PREVIOUS PHASE

Under SEE II, foundations laid from SEE I were reinforced, sustained and built upon through technical advice, capacity building and ICT implementation support. Diagnostics assessments conducted under Phase I, which determined reform trajectory in the beneficiary countries, were repeated under SEE II to take stock of reforms implemented and reset reform priorities.[13] Reforms initiated under SEE I were built on to support the refinement of management and governance structures and arrangements; including the strengthening of core tax functions and the introduction of new ICT systems.

IMF CD support initiated during the previous phase has achieved significant results in a number of areas: i) multi-year reform programs were developed and instituted in all RAs with substantial improvements made subsequently in enhancing governance and management approaches; ii) modern CRM methodologies were initiated to allow for a phased implementation across all countries, currently being implemented; iii) focus was placed on data driven approaches with good investments made by some RAs in the development of data analytics to enhance the identification of major compliance risk clusters; iv) major ICT development projects to replace or upgrade outdated ICT systems across a number of countries, were initiated, that the current phase has started utilising; v) e-tax services such as electronic filing have been instituted by RAs further strengthening tax modernization efforts; and vi) Some key tax functions such as debt collection have been consolidated by some RAs in line with international good practice, and previous phase recommendations.

[13] Results of these recent diagnostics were not available at the time of writing.

4. ASSESSMENT OF PUBLIC FINANCIAL MANAGEMENT WORKSTREAM

4.1. KEY ACHIEVEMENTS

The overall assessment of the Evaluation is that the SEE II PFM reviewed portfolio has performed reasonably well against its targeted outcomes and objectives, despite the adverse context for reform arising from natural disasters, the COVID pandemic, and political instability that affected some beneficiary countries. Key areas of progress under the PFM workstream include:

- The preparation and approval of revised country PFM reform strategies. While all the countries had existing strategies, these had been prepared five or more years earlier. The new strategies provide important updates both at the policy level and in presenting more detailed work plans, which typically reflect very closely key IMF recommendations on the key topics. The high level approval of these strategy documents is an important precondition for accelerated reform implementation.
- There has been good progress in policy development following the completion of PIMA assessments in each of the beneficiary countries. The PIMA assessments, carried out between 2016 and 2020, provide a detailed guide to the gaps in policies and systems for effective management of public investment, including the use of PPP.
- The Program has been successful in providing guidance on the need for legislative reform. Several SEE II countries have adopted policies and taken actions that closely match IMF recommendations. However, the degree of implementation of specific activity plans has been mixed between countries. Progress in PIM/PPP reform has been particularly slow to take effect, mirroring the complex political constraints to any reform of such systems.
- SEE II has supported the strengthening of cash-flow monitoring forecasting and related debt management, a focus of the previous phase when strong foundations in this area were laid. The need for this emphasis arose from the tighter fiscal situation that prevailed in most countries in the region as a result of COVID. Detailed technical recommendations have been provided on the information requirements for managing cash-flow forecasting, typically putting emphasis on the building up of bottom-up information flows on the cash situation of individual budget institutions. Significant improvements have occurred in implementing these strengthened systems in several countries.
- The Program has been successful in getting greater understanding of the need for strengthened systems for the identification, monitoring and management of fiscal risks, which have a potential to disrupt planned fiscal stability. These efforts have concentrated on several risks, including local government budgets, State Owned Enterprises (SOEs), budget arrears, court risks, and VAT refund arrears. SEE II has typically supported the establishment of dedicated Risk Units in the MoF's and in other relevant offices.

4.2. PFM WORKSTREAM-LEVEL ASSESSMENT

PFM workstream				
Overall			Country	Topic
2.7			2.6	2.7
OBJ 1	OBJ 2	OBJ 3	OBJ 4	OBJ 5
3.2	2.5	2.5	3.2	2.8

The overall assessment of this review is that the IMF's PFM portfolio has performed reasonably well against its objectives. The PFM portfolio assessed under the evaluation included five SEE II objectives, which were evaluated through a sample of 15 projects in three country case studies, and one topical case study, which combined PIM and PPP (including PPP fiscal risk assessment model (PFRAM) training), and that was assessed in all six SEE II countries. The average score for the PFM workstream, across all OECD-DAC criteria (comprising all three case study countries and the topical area), for each Objective was: i) ***Objective 1**, PFM Laws and institutions, 3.2*[14] *(i.e., rating of 'good'); ii)* ***Objective 2**, Budget preparation, 2.5 (between 'good' and 'modest'); iii)* ***Objective 3** – Budget execution and control, similarly 2.5; iv)* ***Objective 4** – Asset and liability management (ALM), 3.2 ('good'); and* ***Objective 5** – Identification monitoring and management of fiscal risks, scores 2.8 ('good'). The overall score for the workstream was 2.7 ('good', or 'largely achieved').* The high score for

[14] Evaluation ratings are, 4 = 'excellent'/'fully achieved'; 3 = 'good'/largely achieved'; 2 = 'modest'/partially achieved'; and 1 = 'poor'/'not achieved').

Objective 1 results from there being only a single topic under that objective, that was completed early in the Program.

Country performance across the PFM assessed portfolio based on OECD-DAC criteria showed some variation in achievement, with the overall rating being close to 'good'. Whilst relevance and coherence remained high, the unanticipated advent of the pandemic had an adverse effect on effectiveness, impact and sustainability to different degrees in all three case study countries. This resulted in project interventions being delayed and, in some instances, some milestones required readjustments to their achievement timeline. Some of the delays were mitigated by remote CD delivery, however this incentive was only fully taken advantage of in certain countries where appropriate structures and adequate level of digitalization existed to support remote working. The Average of OECD-DAC criteria scores across all five objectives from the country studies was between 2.3 and 3.5, with the impact and sustainability scores in one country influencing the lower range of average score.

The overall score for the topical case study, at 2.7 ('good'), is marginally higher than the average overall score for the three country assessments. The Evaluation undertook a single topical case study relating to PIM and PPP. Originally planned as two separate topical studies, it was decided to assess them together. This is because PPP is an intrinsic part of wider public investment, and policy recommendations cover the two topics, as for example in the recommendation of a single project pipeline combining budget-funded PIM and PPP projects. The PIM/PPP case study notes the relatively slow progress in these topics across the region, which is seen as springing from the constraints to policy reform arising from political economy considerations. The topics scored particularly highly for coherence, which was awarded an 'excellent' rating.

4.3. RELEVANCE

Relevance				
Overall			Country	Topic
3.4			3.5	3.0
OBJ 1	OBJ 2	OBJ 3	OBJ 4	OBJ 5
4.0	3.0	3.3	4.0	3.5

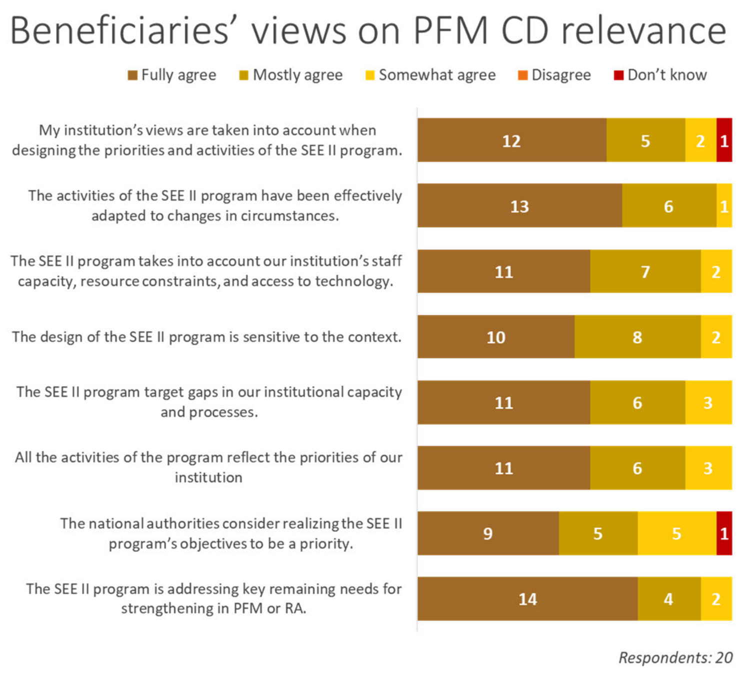
The country case study average relevance scores were between 3.0 and 4.0. This relatively high score reflects IMF's (including the previous SEE phase) continuous support in the region for the past decade and more. This continuity and a large volume of diagnostic and analytical work fed into the design of Phase II. Project design at country level has been identified either through the well-established diagnostic tools, such as the PEFA and PIMA, or IMF experience gained from country case studies. The assessment of relevance also recognises that the workstream has been very directly addressing the main concerns of SEE II partner governments and the EU authorities in the context of the accession process. The scores would have been even higher if political economy considerations were incorporated in project design. This could have been achieved through less over-optimistic implementation schedules which recognised likely implementation impediments arising from political economy issues. For example, under PIM and progress towards unified treatment across PIM and PPP in project selection and entry to the budget can be expected to be slow.

The buy-in from the national authorities has generally been good. This is demonstrated in two ways. First, the authorities have requested each and every intervention undertaken in the Program, indicating at least a good level of awareness of the purpose and likely content of the support. Second, there has been a symbiotic relationship between the research and analysis undertaken through the SEE II TA reports and the shifts in strategy and policy positions of the governments. As an illustration of this relationship, the experience of the countries observed in the field of PIM and PPP management had been raised in the course of the PIMA exercises, most of which were undertaken in 2016 (some years before the start of phase two). This generated the interest for including PIM/PPP in the design of the Program. However, the research analysis and policy work conducted by the missions in this area developed a much more specific agenda than was explicit in the PIMA studies. This greater level of detail in the mission policy proposals have been picked up in the majority of the new round of PFM strategies, which closely reflect the advice on PFM/PPP set out in the mission reports.

KEY FINDINGS FROM THE EVALUATION OF SEE II PFM WORKSTREAM
What has been done well and has been successful
▪ **The Program has responded flexibly to the change in the economic and fiscal environment** arising from physical disasters, political changes, and the COVID pandemic. This has involved a significant reorientation of the priorities of the programme to giving greater attention to improvements in the central government budgetary process and the need to improve the identification of fiscal risks. ▪ **High quality of technical and policy reports.** The quality of mission and technical report on policy options and recommendations for and management procedures to generally under-staffed institutions responsible for advancing the PFM reform agenda is uniformly very high. ▪ **Access to IMF documents and recommendations**. The IMF has ensured access of the key officials to the full mission reports and recommendations. This has facilitated the adoption of key reform recommendations in the revised government strategic documents, in particular the revised PFM strategy. ▪ **The measures taken to overcome the challenges posed by COVID have been well managed.** These have included extensive use of remote delivery both for technical papers and reports, and for the provision of training courses on a remote basis. In terms of technical reports, comprehensive support has been provided at a very detailed technical level on appropriate approaches to the strengthening of the budgetary management system in a stepwise manner. Also the technical support in a number of topical areas displayed a high quality of collaboration between the experts involved and the senior managers responsible for this reform in MoFs. ▪ **The modification of the strategic approach to PFM in one of the case study countries in particular was well-managed.** Specifically the recognition that the emphasis of the country's SEE II program needed to be adjusted to take adequate account of the increased fiscal risks posed by the fiscal deterioration emanating from the COVID pandemic was appropriate, as was the specific focus of support on the risks posed by the SOE sector.
Areas for potential improvement
▪ **There is a need for increased political economy analysis** to be undertaken especially for reform areas that are expected to encounter challenges from different stakeholders with different interest in areas such as PIM and PPP reform. This analysis should concentrate on identifying the structural aspects of the country political and government system. Giving appropriate attention to the political economy would likely increase the ability of the project to nuance the approach to complex areas such as PIM/PPP reform. This should also take account of the position of the country in relation to EU accession which is expected to be the principle driver for actual reform implementation in the coming years. ▪ **Enter written agreements on reform support with the government authorities**: A more formal approach to agreement with the government authorities on the scope of the program, the outcomes and milestones to be targeted, the roles of the various stakeholder agencies in delivering the reforms and the complementary assistance required from other development partners. This should take the form of a written agreement between SEE II/IMF and the national authorities. ▪ **Strengthen the strategic relationship with the EU.** The Program should seek to increase the depth of the relationship between the IMF and especially the EU, which has the presence on the ground, the scale of operations and the access to the high political level in government to be able to contribute substantially to the effectiveness and outcome of the project. SECO's strong presence in most SEE II countries can also be leveraged further. ▪ **Access to long-term CD support**: Consideration should be given to increasing the availability of longer-term project-type TA to provide strengthened follow-up to PFM policy recommendations adopted by the government, either through SEE II, or in collaboration with development partners. ▪ **Attention needs to be paid to build greater awareness in government of the specific objectives and milestones to achieve a common understanding across the IMF/beneficiary divide**. The evaluation noted that in almost all cases, the senior officials responsible in specific topic areas are not aware of the milestones set by the IMF and reported on in the IMF RBM system. ▪ **Mechanisms for informing the development partners of progress on each intervention could be strengthened in some countries.** Although channels exist including the meetings of PFM donor Committees these should be strengthened to provide more up to date monitoring information.

The Program has successfully adapted to new and emerging capacity gaps. This is illustrated by the major change in emphasis and direction that occurred following the onset of COVID. Specifically some of the Program areas of lower consequence in the new circumstances were dropped or delayed, while new areas of concentration were introduced. In particular, in direct response to the effects of COVID on fiscal stability and tightening, the SEE II design was modified strongly to cover newly identified weaknesses: in particular cash and debt management and the identification, monitoring and management of fiscal risks. In most countries several new analytical units have been newly set up or expanded under the revisions to the Program.

Political economy considerations may have been better reflected in the design of PIM and PPP projects, and this factor lowers the relevance score of the case study to 3.0. PIM is particularly exposed to delays in the implementation of reforms, owing to strong vested interests at in the support for adoption of particular projects based on political motivations. The TA reports on PIM show inadequate understating of this crucial role of political interests and this results in over-ambitious targets for the achievement of technically straightforward, but politically complex reform. The lower score for Objective 2 arises from a country case in which the evaluation considers that the inclusion of a specific topic was questionable on grounds of prioritisation, in the view of past lack of buy-in on the part of the authorities.



PFM CD beneficiaries' survey responses confirm the relatively high relevance of the Program. A significant proportion of respondents feel their views are taken into account, albeit less strongly than in RA; that CD is addressing key priorities (more strongly than RA); and that delivery has been responsive (similar to RA). However, a fewer proportion fully agreed that the Program is sensitive to the context; or that national authorities share in the institution's prioritisation of SEE II activities (similar concerns as RA).

4.4. Coherence

Coherence				
Overall			Country	Topic
3.2			3.0	4.0
OBJ 1	OBJ 2	OBJ 3	OBJ 4	OBJ 5
4.0	3.0	3.3	3.0	3.3

The overall score for coherence was 3.2, i.e., rating of 'good'. SEE II and its previous phase were built on an unusually close collaborative understanding of the priorities for PFM reform. A key feature of the Program is the underlying complementarity of the agencies. The EU has brought momentum for the Program in creating strong motivation linked to substantial financing options available on a grant basis. The IMF brings its high level of experience and understanding in the specific topics under PFM (and reinforcement through the surveillance process), while the World Bank has brought continuity in country and funding instruments to support long-term capacity building. The view that SEE II has involved an unusual degree of complementarity and synergy between these agencies was expressed in many interviews with beneficiary representatives. One survey response, however, highlighted a lack of coordination in the implementation stage between IMF and a DP in one of the SEE II countries.

The close coherence between the SEE II and other PFM activities in related fields was strengthened by the participation of the major DPs in the drafting of country development plans, which fed directly into the PFM reform strategies. In one country the DPs sat in the government committee which defined the progression from the original PFM strategy to the revised version. Strong synergy with the wider IMF fiscal policy aims is evident in the extent of commonality in the findings and recommendations of IMF Article IV missions and SEE II workplans. Alignment was achieved with other agencies involved in PFM, including GIZ and USAID. In one country GIZ and IMF worked together on the development of a budget manual.

There were variations in coherence ratings between countries and objectives. The country assessments scored coherence at an average of 3.0 ('good', or 'largely achieved'), with similar scores achieved by all countries. There is significant variation in the scores for coherence across objectives, with a range from 3.0 to 4.0. In a context of almost uniformly high scores for the coherence criterion, there is actually a single outlier with a low score in a country where political instability led to low levels of overall SEE II activity. In the one topical case study of PIM and PPP, the coherence criterion scored a perfect 4.0 ('excellent').

The PFM CD beneficiaries survey findings confirm the Evaluation's coherence-related assessments based on other information sources. 80% fully agreed that CD recommendations are consistent with other IMF advice; 75% that they are consistent with, and help the country's broader economic program and EU accession agenda; while 60% that CD is effectively coordinated with other DPs. The balance was split equally between mostly agree and somewhat agree in respect of the first two questions, while for the third question, 12% mostly agreed, and 20% somewhat agreed.

4.5. Effectiveness

Effectiveness				
Overall			Country	Topic
2.4			2.4	2.3
OBJ 1	OBJ 2	OBJ 3	OBJ 4	OBJ 5
2.0	2.3	2.3	3.0	2.8

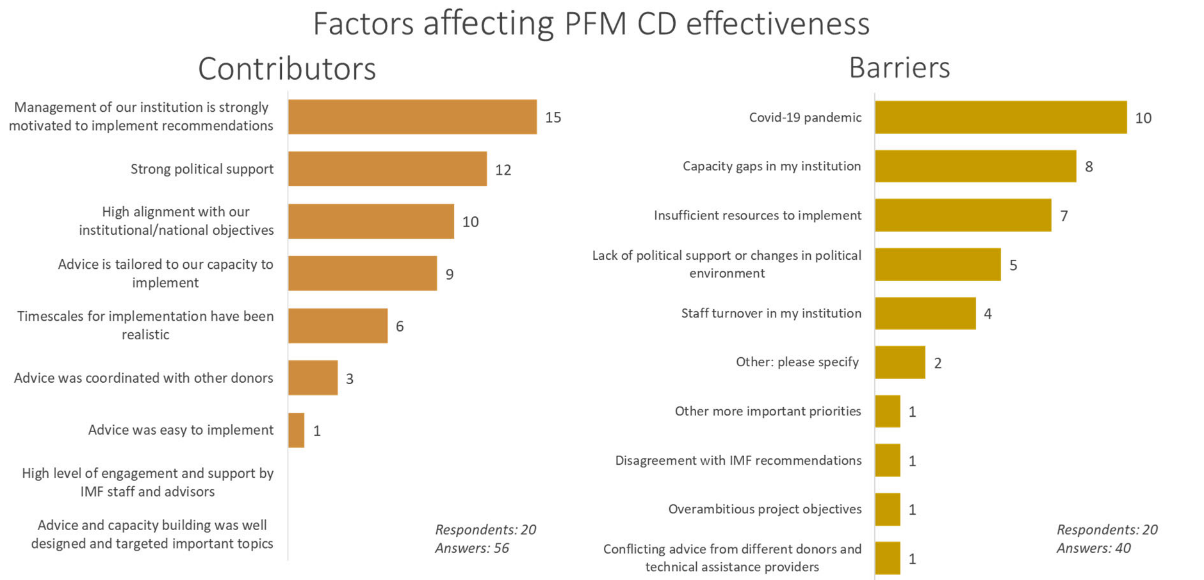
The PFM workstream scored an average of 2.4 (rated between 'good' and 'modest') for the effectiveness criterion, and across the three country case studies. While in most cases the evaluation endorsed the IMF's own assessment of scores at the outcome and milestone levels, in a few cases (mainly upward) amendments were suggested. These mostly related to taking on board the progress which had been made in the period since the last assessment by IMF.

The relatively low PFM score on effectiveness (and to a degree efficiency) reflects constraints arising from the COVID pandemic and, in selected countries, political instability. With several countries under lockdown for an extended period, normal office working was seriously affected, and the adoption of remote working provided only a partial substitute for in-person missions and interactions. The reduced delivery of reform implementation outcomes and milestone in the review period was to a degree offset by good process at the level of strategic planning for PFM reform. In several countries important PFM strategy documents were revised or written anew.

Another factor affecting effectiveness and efficiency of the PFM workstream in particular is the need to work across different government entities. PFM reforms are inherently complex, and need buy-n from multiple departments and ministries. In areas such as PIM and PPP, additional complications arise as often local governments need to get involved.

The overriding factor leading to depressed scores for effectiveness as measured by milestone and outcome achievement is the disruptive impact of the pandemic. Effectiveness of SEE II has been supported, if not fully, during the COVID-19 period by two main factors: i) continuing impetus for reform at the strategic level in the beneficiary governments arising from the EU accession process. This factor has manifested itself in the progress made in preparing, and approving key strategy documents in several countries; and ii) progressing with institutional developments including the establishment of policy units in areas such as cash flow, debt management, SOE oversight and fiscal risk. In some countries this progress has been significant, but in most progress has been tempered by difficulties in recruiting and retraining appropriately trained and experienced staff. The adoption of remote operation by the IMF has also helped to sustain progress as has the shift of program emphasis to topics to address new fiscal concerns.

The range of effectiveness scores for individual countries ranged between 2.3 and 2.6, while the topical case study on PIM and PPP scored 2.3 (high 'partially achieved') for effectiveness. In general, there is only limited variation in effectiveness scores both between and within countries. The explanation for this lack of variation is believed to be the suppressive role of the COVID pandemic, which led to low achievement of outcome and milestone targets. Similar considerations apply to the limited variation between objectives within countries and the topical case study in PIM/PPP. The latter was also disproportionately affected by delays in decision-making at political level. It has been observed by the Evaluation that officials in the region are reluctant to take responsibility for reform implementation until there is unequivocal support from a high level of government.



Surveyed beneficiaries attribute the success in CD implementation mainly to motivated management and political support. In contrast to the RA survey, the role of IMF was not prominent. IMF staff and LTX survey, on the other hand, attributed effectiveness principally to the Fund's engagement, and well designed and targeted CD. Management motivation and political support were ranked third and fourth in the order of importance. Beneficiaries highlighted COVID-19 as the main barrier to implementation, followed by capacity gaps and lack of resources. In contrast, IMF staff and LTX ranked lack of political support and changes in political environment, staff turnover, and capacity gaps as the three most important impediments. These differences in perspectives between IMF and the beneficiaries should be explored further in Phase III to identify and address the actual success and failure factors.

4.6. Efficiency

Efficiency scored an average of 2.3 across the three country case studies, and 2.0 on the topical case study, while the overall average assessment on all projects for the workstream was 2.3 (rated between 'good' and 'modest').

Efficiency				
Overall			Country	Topic
2.3			2.3	2.0
OBJ 1	OBJ 2	OBJ 3	OBJ 4	OBJ 5
3.0	2.5	2.3	3.0	2.3

The assessment of efficiency was influenced in all country and the topical case study for PIM/PPP studies by two offsetting factors. On the one hand COVID has had the effect of reduced direct costs compared to budget as in-country missions were discontinued early in 2020, which rendered actual costs under SEE-II non-comparable to the costs incurred under Phase 1. The discontinuation of in-country missions is believed conversely to have led to a perceived but unquantifiable reduction in the value of the benefits accruing from the Program. As highlighted under effectiveness, this reduction in benefits was considerable and took the form of weaker than expected achievement of outcomes and milestones.

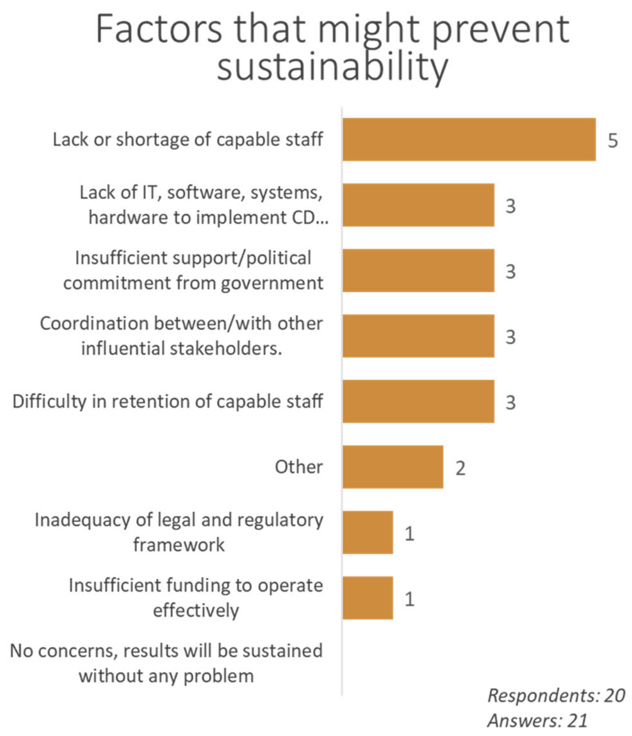
The assessment of SEE II costs incurred relative to alternative delivery systems is difficult to determine. A rigorous assessment would require the definition of counterfactual alternative system(s) of project delivery, which is beyond the scope of this evaluation. However, it can be argued that greater efficiency and effectiveness might be achieved through increasing the scope of follow-up support at the reform implementation stage, identified as a weak point in existing project delivery. Moreover, cost assessment is difficult as insufficient breakdown of costs of project activities produced by the IMF financial information systems (CDROM), which was operational during the period under review. It is understood that the current project management information system, CDMAP, provides more detailed cost breakdowns.

CD beneficiaries value all modes of delivery, with that from HQ staff being considered marginally more effective. In the survey conducted as part of the Evaluation, IMF HQ staff were ranked the highest with respect to the effectiveness of CD delivery channels, followed marginally behind by LTX, and the STX.

4.7. IMPACT AND SUSTAINABILITY

Impact				
Overall			Country	Topic
2.3			2.3	2.0
OBJ 1	OBJ 2	OBJ 3	OBJ 4	OBJ 5
3.0	2.3	2.0	3.0	2.3

The scoring for impact and sustainability are treated together as they are closely linked, since impact lays the basis for sustainability. The impact criterion was scored an average of 2.3 across the country case studies and 2.0 in the topical case study. The overall average score for impact for the workstream was 2.3 (between 'good' and 'modest'). Variation between countries fell in a range of 2.8 and 1.3, with one country in particular being adversely affected for being a low recipient of CD. The sustainability criterion scored an average of 2.3 across the countries, with an overall average including the topical case study of 2.4.



Source: SEE II evaluation CD beneficiaries survey

The assessments of impact and sustainability each cast a note of optimism, combined with one of caution. It might be expected that the relatively low scores for effectiveness, which are based on actual progress at the outcome and milestone levels, would negatively affect impact and sustainability. While this is likely in the short-term, looking at the medium-term, all the selected case study countries have been paying attention to strengthening their strategy and policy frameworks for continued PFM reform progress, despite disruptions caused by COVID and, in some countries, political instability. This is seen as an important prerequisite for recovery and acceleration of reform implementation as the COVID threat recedes.

However, two caveats remain: first the resumption and acceleration of reform activity has yet to be strongly manifest, although the potential certainly exists and has been attested to in several beneficiary interviews. Second, as the technical demands of the Program, including major requirements in terms of institutional and individual capacity to staff in the wide range of technical units which have been created, continues to cast a shadow on the prospects for accelerated reform implementation. As a result of these considerations the evaluation has taken a conservative approach in scoring impact and sustainability.

Sustainability				
Overall			Country	Topic
2.4			2.3	2.7
OBJ 1	OBJ 2	OBJ 3	OBJ 4	OBJ 5
3.0	2.3	2.3	3.0	2.5

With the exception of Objectives 1 (laws and institutions) and 4, ALM, the scores for the impact and sustainability DAC criteria are rather even across and within countries. The high score for Objective 1 results from the fact that this objective was only represented by a single topic in a single country where the full results of the intervention were realised early in the Phase 2 of the Program and were embedded in strategic documents and institutional innovations that have already been put in place. The evaluation takes a positive view of the future prospects for sustainability, based on the groundwork already achieved in adoption of strategic and detailed policies in key government documents, and as a result of momentum from the EU accession process, which provides strong incentives to continue existing reforms. However, this rosy view is questioned to a degree by the issues in all countries relating to the difficulty in staffing the new technical PFM units currently under implementation. In light of these considerations the evaluation takes a relatively conservative view on sustainability at the present.

4.8. Outcomes from PFM CD delivered in the previous phase

The SEE II programme built on and developed the previous SEE I project which ran from December 2015 to April 2018. SEE I itself built on programs of both the IMF and EC from an earlier period. Specific achievements of the SEE-I project included:

- The development of PFM reform strategies in five of the six CD beneficiary countries.
- Strengthening of annual budget processes and the introduction on medium-term budget frameworks.
- Strengthening of the legal basis for PFM including preparation and in two countries, the passage of Organic Budget Laws.
- Foundational analysis of public investment including the completion of PIMA assessments in four countries,
- Work on fiscal risk management and cash flow forecasting and management.

SEE II has built on these SEE I activities with a substantial strengthening of support in the areas of PIM and PPP, fiscal risk and cash flow management, guided strongly by new priorities emerging under the COVID pandemic and its impact of fiscal management and sustainability, which were the subject of review and analysis in mission reports prepared at the onset of the pandemic.

5. REVIEW OF SEE II STRATEGY AND OPERATIONS MANAGEMENT

5.1. ADAPTABILITY TO EMERGING ISSUES

SEE II adapted well to the environment that resulted from the break-out of COVID-19, both operationally, and strategically. At the onset of the pandemic, CD delivery missions and events needed to be paused or rescheduled. These were moved online relatively rapidly, and, for example, FY2021 CD for both workstreams was broadly in line with plans. Strategically, the Program expeditiously refocused to meeting counterpart emerging needs. A number of COVID-related guidance papers were produced at the onset, for example on RA on customs administration responses to the emergent risks; business continuity of RAs; and operations to safeguard collections. In PFM, similar guidance was produced on managing fiscal risks and expenditure arising from, and managing PFM systems impacted by, COVID-19. Over time, again partially to meet counterpart demand, the emphases of topics covered was adapted. In PFM for example, managing the fiscal (contingent) risk arising from government COVID support measures, as well and cash and debt management increased in focus. The RA workstream, impact areas for preserving or strengthening revenue collection, such as management of Arrears and CRM, saw more activity. Although well mitigated, the adverse impacts of COVID-19 have not been fully alleviated. In the Evaluation's surveys of CD beneficiaries' and IMF staff and LTX for example, 92% of IMF and 89% of beneficiaries indicated that remote delivery was between highly and somewhat less effective than in-person missions. In addition to absorption of CD, more limited implementation guidance on more complex technical areas, and the lack of peer-to-peer interactions afforded by regional workshops, staff shortages at beneficiary institutions and the need to focus on priorities arising from COVID-19 will have slowed implementation. Absence of physical travel, for example by LTX, meant that it was more difficult to monitor the Program and maintain dialogue with counterparts. Going forward, hybrid CD delivery models are clearly going to be warranted, and a key task will be to identify the appropriate tools for different types of CD, through piloting, and retaining flexibility.

More limited work has been done in cross-cutting themes such as climate change, gender, and for example, corruption. The RA workstream includes areas such as excise duties, that have climate impact. Going forward, there are plans to introduce the PIMA climate change module. On gender, the proportion of females attending workshops are recorded, but not reported on. IMF Governance missions in coordination with Area Department mission chiefs advise on corruption, and communications are maintained to keep an overview of the situation. To date, therefore, addressing these areas have not been prioritised to pro-actively target or tailor interventions strategically or for individual projects, for example in PIM (climate change). Behavioural approaches, for example in tax collection, or large taxpayer office and high wealth individuals, could be additional cross-cutting areas for consideration. In order to embed these cross-cutting themes into SEE III, action needs to be taken at both the workstream design level (to identify topical areas where positive impacts made, or negative impacts mitigated), and in the design of individual interventions, so that policies, regulations, and rules can be appropriately designed. HR and operational policies in counterpart institutions can also be covered. It should be recognised, nevertheless, that these cross-cutting themes (especially gender), are not often priorities of counterpart institutions, and care needs to be taken to minimise frictions with their relationships with IMF, by taking a flexible and opportunistic approach.

5.2. LESSONS FROM AND ADDRESSING SLOW IMPLEMENTATION

Slower than planned implementation has been caused by a number of factors, all beyond IMF control. These impediments are highlighted in more detail in Sections 3 and 4, on how they affected the PFM and RA workstreams respectively. The key reasons are: i) COVID-19; ii) political instability and changes in government leading to delay in decisions at legislative and policy levels, and feeding into administrations level as uncertainty results in key strategic decisions being deferred, and there is rotation of senior managers; iii) staff capacity constraints, under resourcing, and high staff turnover in counterpart institutions and functions; iv) delays in implementation of key ICT systems, such as the Integrated Financial Management Information System (IFMIS), or core revenue platforms; and v) unrealistic implementation schedules given the above constraints and complexity of many proposed reforms, giving an impression of delays.

While IMF has made significant progress in recent years in shifting focus from outputs to outcomes, and the RBM logframe system has helped in this regard, further work is needed in developing tools to identify and address actual and potential constraints in achieving results, and monitor implementation. For example, although political risks and capacity constraints, which would have a big impact on translating outputs to outcomes through implementation, are assessed through degrees of formality at the stage of program design, and monitored informally during implementation, the process was not systematised during the period under review, but is now understood to have been. Moreover, political economy affects implementation of the Program as a whole, but different dynamics can affect particular topics, such as PIM, or RAs implementing HWI compliance programs. More systematically embedded political economy analyses in the program activities (at least in the higher-risk areas) would enable anticipation of potential blockages and interventions designed (for example, to include additional stakeholders), or sequenced (to ensure impediments are addressed at the appropriate time) to at least mitigate potential barriers to implementation. A detailed institutional (including leadership, motivation, and change management) capacity gap analysis would similarly identify in advance situations where lack of resources or capacity is a constraint, and the project design (for example, more intensive CD delivery, additional training, bringing in DPs to provide long-term capacity building support) should be designed in, and implementation timescale adjusted appropriately. In the latter case, counterparts could also be requested more formally to ensure adequate budgets and/or staffing to ensure implementation is effective, although IMF has pointed out that it has limited leverage in this regard. More focus on sequencing interventions, for example when implementation of a project's recommendation has dependencies, for example the implementation of an IT system, might reduce delays in some cases. The RBM system is project-based, and linear, defining, and recording activities, outputs and outcomes, and does not capture the risks and assumptions that a traditional logframe would do. It does not capture linkages and interdependencies between projects, so is not intended to aid sequencing. Moreover, the IMF CD delivery approach, principally missions-based, does not lend to regular monitoring of implementation progress. Although LTXs are the best placed to do this task, this has not been the standard, and all missions, including those covered by LTXs, have to date typically focused on the topic(s) covered in their mission terms of references. In most cases, milestones are updated annually, at the time of the annual review, or for specific projects when a related mission on that topic is undertaken. More regular, systematic monitoring of implementation of all projects, at least quarterly, would help identify actual or potential risks of slowdown, and appropriate, pro-active action taken. These steps would not necessarily prevent all delays, especially from political causes, but at least mitigate impact on program implementation and efficiency. IMF has pointed out that staff have other responsibilities that does not allow monitoring of all recommendations, and LTXs have to maintain a balance between monitoring previous advice and new engagements.

5.3. Partner visibility and implementation of SEE II operational guidelines

There is scope for improvement in actions to enhance Partner visibility. Actions relating to visibility have been mainly focused on some social media postings, and displaying donor logos on workshop slides and selected outputs. This has had limited impact to date. For example, of the workshop participants surveyed, of nine respondents one was aware of all of the donors, four some, and four none. Out of seven respondents from EC and SECO country offices to a question of satisfaction with visibility, two fully agreed, none mostly agreed, two somewhat agreed, and three disagreed. The local Partner offices feel that among beneficiaries, SEE II is seen as an IMF program, with limited or no awareness of SECO and EC as donors. A number of EC and SECO offices have suggested joint briefing meetings with senior counterparts (which happen occasionally) to not only improve visibility, but also strengthen the common messages being delivered by important development partners.

The administrative provisions of the SEE II operational guidelines appear to be broadly followed, but implementation could be improved in selected areas. According to the guidelines, one of the steering committee's role is to provide strategic guidance. Feedback from interviewees suggests that meetings are mainly focused on querying issues raised in the progress reports and discussing work plans. Including in the meeting agenda a set of strategic questions for discussion may refocus steering committee meetings. Some country offices indicated that they do not get adequate advance notice (or not at all) of IMF missions (while at the same time IMF has highlighted country offices do not provide information on work in related areas to the Fund). Some local offices cannot access the Partner platform for documents, but that their headquarters download and share them. Delays in access to TA reports is a continuing issue, but some LTXs and HQ staff provide verbal debriefings of key points (the experience of interaction between the SECO and EC delegations tends to vary between workstream (more interviewees being satisfied with RA), and the LTX in post. Under the operational guidelines, EU and SECO are expected to share documentation related to CD they are planning or providing in related areas in the region. This is understood to not take place. Overall, EU and SECO country offices are generally satisfied with their cooperation with IMF under SEE II. In our survey, out of 10 respondents four were very satisfied, five where mostly satisfied, none somewhat satisfied, and one mostly dissatisfied (none highly dissatisfied).

5.4. SEE II STRATEGIC ISSUES

The SEE II strategic approach appears to be based on solid technical diagnostics, overlayed by careful consideration of beneficiary demand. In both the RA and PFM workstreams, diagnostics such as TADAT, PEFA and PIMA have played a major role in identifying gaps that need to be addressed to strengthen effectiveness. These diagnostics have provided a robust underpinning to Program's activities, and in achieving, or progressing results for individual projects. Responding to recipient demand, or changes in the environment (for example with respect to COVID-19) provide flexibility in implementation of projects. The focus appears to be on achieving results in individual interventions, with the assumption that once they are all achieved, higher goals (for example, tax-to-GDP ratio, or measures for fiscal discipline) would be reached. Given SEE II can't implement all the actions needed to reach these higher objectives, and not even many of the more important ones, there is an implicit reliance on other DPs supporting areas not covered by the program, and that the counterpart institution's own strategy is aligned with Program activities. In SEE II countries, the Accession requirements provide a strategic framework for this purpose.

SEE II therefore follows more of a bottom-up strategic approach. The RBM logframe reinforces this process, with focus on being achieving individual projects at the outcome level. There is no explicit linkage between the achievement of individual outcomes and their related objectives (for example, more improved management and governance in RA), and the dependencies between the objectives (for example, the achievement of the two RA objectives are inter-dependent to a large extent, and delays in one can impact on implementation of projects in the other). In any event, the RBM objectives, as defined, are essentially higher level outcomes, and no measures (and therefore targets) are provided in the logframe for higher-level impacts, or to demonstrate how the RBM objectives would contribute to them. A more top-down strategy, starting from defining the higher-level goals, would identify what needs to be done to achieve them; who will do it; the appropriate sequencing needed; and what role an individual institution or donor program will play within a set of mutually supporting activities. This gap between the current bottom-up bias of the Program, and developing a more explicit linkage with higher-level strategic goals (in order to help prioritise activities, identify important gaps not being addressed, and identifying and addressing key risks both to the Program and those affecting areas not covered by it) could be addressed by developing a robust and comprehensive Theory of Change.

6. SEE II WORKSHOPS AND WEBINARS

Eight SEE II workshops and webinars were delivered during the review period, five by the PFM workstream, and three by RA. PFM events covered PIMA and related infrastructure governance topics (three); implementation measures to address COVID-19 (one); and fiscal risks from SOEs (one). Topics covered by RA events were CRM, IT, and VAT Gap. Due to COVID-19, five of the workshops were delivered remotely.

The workshops and webinars are considered by the Evaluation to be of high quality. Workshops have been evaluated as part of the projects in the topics covered in CD assessments. A sample of presentation outputs was also reviewed, and they are assessed to be of high quality, and practical, giving implementable advice. Post-workshop participant feedback forms are not canvassed by RA. For PFM, feedback from two workshops (on SOE fiscal risks and cash management) was available for review, and both scored very highly in participant satisfaction (minimum 4 out of 5, with many scores ranging between 4.6 and 4.9) . A survey was conducted of workshop participants as part of the Evaluation. Only the contact lists for four workshops dating from 2019 and 2020 were available. Only nine (out of 96 contacted) responses were received. The main workshop benefits highlighted by participants were: becoming aware of international best practice in the topic; gaining practical, implementable knowledge; and developing technical knowledge. More than half the respondents indicated that they had implemented almost 75% of the lessons learnt during the workshops. A key benefit of the workshops is seen to be the opportunity of peer-to-peer learning and networking, and this was missed during the webinars.

SEE II workshops and seminars are closely coordinated with and are integrated with the Program's other CD activities. Workshop topics covered in PFM, e.g., PIMA and infrastructure governance, and RA, such as CRM and IT, have complemented core technical advisory activities. Both workstreams also organised workshops on pertinent COVID-19-related topics at the onset of the pandemic, which, in addition to disseminating technical advice, provided an important opportunity for peer-to-peer learning. In the survey of workshop participants, all respondents rated the linkage between SEE II workshop contents and technical advice as being excellent (the highest rating).

7. MAIN FINDINGS AND RECOMMENDATIONS

The following are the principal findings and recommendations arising from the Evaluation.

	Findings	Recommendations	Priority
1.	Political risk is high in the region. PFM implementation, e.g., in PIM, has been slow due to political barriers. In RA, in addition to lack of senior endorsement of changes in one country, implementation of initiatives such as CRM will likely face resistance from vested interests.	**Undertaking formal analyses of political, governance, corruption and institutional systems and structures in higher risk SEE II countries, at both Program and project levels, to identify and address potential barriers to effective implementation of CD:** Consider, as part of designing SEE III, undertaking formal reviews covering the structural aspects of the political, governance and institutional systems in high risk SEE countries and the associated vested interests that may impede reform in PFM and RA. In addition to program-level review, consider assessing related risks associated with implementation of particularly sensitive topical areas. Preferably, such reviews should be systematically updated periodically, say six-monthly. They will potentially enable more proactive action (e.g., escalation, re-sequencing, bringing into discussions particular stakeholder groups, stop/go) to be taken. Special attention should be paid to addressing the primarily political economy constraints to PIM/PPP reform using a savvy PE-informed approach that aims to take incremental steps to implementing the basic revised framework for PIM/PPP reform. This needs to be based on close engagement at the political level to ensure that each required step has high level political buy-in. Where IMF *modus operandi* precludes such engagement, consideration should be given to working with Partners.	High
2.	Counterpart absorptive capacity, both in technical terms and with respect to staffing; lack of technical knowledge of senior staff (especially in RA); and need for major cultural change (also RA) has been a barrier to effective implementation of CD support in some countries.	**Consider undertaking formal Resource Analyses of counterpart staffing levels, skills and motivation to identify and address gaps:** More attention might be given to areas such as leadership, change management, organisational capacity, and institutional infrastructure (esp. ICT, if needed and/or how it might impact on CD), that underpin and enable successful technical reforms. Softer qualitative issues, such as incentives, benefits, knowledge, attitude, skills and behaviours should be considered, especially in situations requiring major changes. Also consider getting counterpart commitment to fill any resource gaps identified as part of the CD agreements being proposed for the next phase.	High
3.	In the PFM workstream robust and comprehensive strategic plans have been approved, and major reform implementations are needed in a range of areas. In RA, important organisational transformation projects have been initiated, which need to be embedded and can take significant time.	**Shift focus more to consolidation and capacity building in SEE III:** Our suggestion is that during the next phase significantly more focus should be on building staff implementation capacity through, e.g., training and hands-on support; and embedding processes. Engaging more long-term resident experts may be necessary in some cases. We understand that the SEE II CD delivery model may not be fully tailored to meet all needs in this area, so proactive action may be needed to engage development partners, especially the EU, in this effort. A differentiated approach may be needed between countries that have stronger capacity and motivation to implement reform, and others that have not.	High

	Findings	Recommendations	Priority
4.	In RA, implementation of CRM requires significant changes in organisational structures, approach and culture, but also opens new opportunities for increasing compliance.	**Explore opportunities for optimizing the potential offered by CRM**: In addition to using CRM to identify risk sectors and audit case selection, and building on the risk differentiation management for LTO taxpayers, expand use of CRM to drive RA activities in supporting voluntary compliance amongst other taxpayers, through awareness, education and outreach programs (which currently do not form a distinct element within the program). This may be an area for coordinated EU (or other donor) support.	Medium
5.	In RA, low data analytics capacity inhibits wider application of CRM, Tax Arrears, and HWI reforms. There is limited matching of data analytics CD to CRM CD, with data analytics supporting CRM in only 1 of 5 countries.	**Strengthen SEE impact in CRM, Tax Arrears and HWI with Data Analytics CD:** Identify data analytics needs, and coordinate/sequence data analytics CD with CRM, tax arrears and HWI CD to maximise the impact and effectiveness of CRM. Coordinate with other donors as data analytics CD may require procurement support.	Medium
6.	A strong motivation of reform in all countries is Accession and the EU has potentially significant influence and leverage, especially in PFM.	**IMF and EU should consider a closer, more strategic partnership for SEE III.** Closer cooperation could extend to a more integrated dialogue with authorities; political escalation where necessary; and coordinated approach to CD interventions (e.g., when longer-term embedded CD-delivery is required). SECO can also play an important partnership role, especially in RA.	Medium
7.	Timescales for implementation of some proposed action plans may not have been realistic given counterpart capacity and other barriers. In addition, in some cases sequencing of activities may not have been optimal (e.g., timing of CD delivery where implementation of IT systems, or staffing up of implementation departments, is a prerequisite).	**Consider developing milestones and outcomes in action plans more based on historic experience and counterpart capacity. More tailored sequencing of CD delivery may be required in some cases.** We understand that as part of the TA agreements with counterparts for SEE III, they will more formally sign-off on action plans including the timetable for implementation. There is potentially some risk in them feeling implicit pressure in agreeing to unrealistic deadlines, including for internal organizational or political reasons. More care may need to be taken by IMF to establish realistic deadlines (some operational changes take a long time to embed), whilst ensuring momentum for implementation is maintained. In addition, the timing of CD delivery in light of dependencies may need more attention. This action may need to be complemented by a more structured monitoring of implementation of milestones (e.g., quarterly).	Medium
8.	Corruption is a major cause of leakage in the region; climate change is an increasingly important issue; and gender is an important cross-cutting theme for donors.	**Consider as part of SEE III engaging experts in anti-corruption (integrity), climate change, and if appropriate gender to advise on embedding these issues into project and CD design.** Such experts, who need to be familiar with the PFM/RA thematic areas, could review the broad areas proposed for SEE III and highlight potential areas where anti-corruption, climate friendly, or gender supportive action may be taken (or there is risk of detriment). They may also provide advice at the project level when CD advice is formulated, when appropriate impactful provisions may be incorporated. It is acknowledged that governments may not be very receptive in this area, so selectivity may be required.	Low