



Independent Evaluation Office
of the International Monetary Fund

BACKGROUND PAPER



BP/15/01

Review of Ex Post Assessments of Countries with Longer-Term Program Engagement and of Ex Post Evaluations of Exceptional Access Arrangements

David Goldsbrough

IEO Background Paper
Independent Evaluation Office
of the International Monetary Fund

**Review of Ex Post Assessments of Countries with
Longer-Term Program Engagement and of
Ex Post Evaluations of Exceptional Access Arrangements**

Prepared by David Goldsbrough

August 26, 2015

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

Author's E-Mail Address: davidgoldsbrough@mac.com

Contents	Page
Abbreviations.....	iv
Foreword.....	v
I. Origins, Objectives, and Coverage of Country-Specific Self-Evaluations	1
II. Methodology of This Review.....	3
III. Effectiveness of the Self-Evaluations	5
A. How Well Do the Self-Evaluations Cover Different Aspects of the Program Engagement?	5
B. EPE- and EPA-Specific Issues.....	13
C. Quality of Different Types of Lessons.....	17
D. Views of Interviewees on the EPA and EPE Processes.....	20
IV. Systems for Following Up on Lessons Arising from Self-Evaluations.....	22
V. Conclusions.....	23
Figures	
1. Content of Discussion of Program Involvement and Program Design.....	7
2. Effectiveness of IMF Involvement	9
3. Forward-Looking Strategy.....	10
4. Overall Quality of Self-Evaluation Reports.....	12
5. Clarity of Lessons and Quality of Overall Report by Type of Team Leader and Type of Country	14
6. Content of Discussion of EPE-Specific Issues	15
7. Content of Discussion of EPA-Specific Issues.....	16
8. Content of Discussion in EPA Updates	17
9. Quality of Different Types of Lessons in Self-Evaluation Reports.....	19
Tables	
1. Self-Assessment of IMF Programs: Coverage, 2003–12.....	3
2. EPAs and EPEs Covered by this Review	4
Annexes	
1. Country Assessment Matrix for Review of EPAs and EPEs.....	25
2. Program Lending Self-Assessments, by Country, 2003–14	27

ABBREVIATIONS

ECF	extended credit facility
EPA	Ex Post Assessment
EPE	Ex Post Evaluation
FCL	Flexible Credit Line
GRA	General Resources Account
HIPC	Heavily Indebted Poor Countries
LTPE	longer-term program engagement
PIN	Public Information Notice
SPR	Strategy, Policy and Review Department (IMF)
PRGF	Poverty Reduction and Growth Facility

Foreword

In April 2015, after this paper was completed, the Executive Board of the IMF decided to repeal the policy requiring Ex Post Assessments (EPAs) for members considered as having Long-Term Program Engagement with the IMF. In place of EPAs, the Board decided that lessons will continue to be drawn from past programs through peer-reviewed assessments to help design successor programs. The experience with EPAs during 2005–13 examined in this Background Paper could inform the design of the modalities for this new approach, which still remain to be decided.

The requirement for Ex Post Evaluations of arrangements providing exceptional access to IMF resources remains unchanged.

I. ORIGINS, OBJECTIVES, AND COVERAGE OF COUNTRY-SPECIFIC SELF-EVALUATIONS

1. This review assesses recent experience with two types of country-specific self-evaluations undertaken within the IMF: Ex Post Assessments (EPAs) of Longer-Term Program Engagements (LTPEs) and Ex Post Evaluations (EPEs) of Exceptional Access Arrangements. While the origins and detailed objectives of the two exercises differ, they both follow the same procedures that are designed to provide a fresh look at the effectiveness of the Fund's involvement with member countries engaged in particular types of arrangements and to draw relevant lessons.

2. The system of EPAs was established in 2003 in response to the report of the Independent Evaluation Office (IEO) on prolonged use of Fund resources. Following the evaluation, the IMF Board endorsed specific measures proposed by a staff task force to help guard against problems that might arise from prolonged use. These measures included new procedures to strengthen the Fund's "due diligence" for countries with LTPE through systematic ex post assessments.¹ EPAs are expected to provide "an analysis of the economic problems facing the country, a critical and frank review of progress during the period of Fund-supported programs, and a forward-looking assessment that takes into account lessons learned and presents a strategy for future Fund engagement."²

3. The IMF Board agreed to require evaluations by the staff of programs supported by exceptional access at the time of its 2002 discussion of Access Policy in Capital Account Crises.³ The aim of such an EPE is to determine whether justifications presented at the outset of a program were consistent with Fund policies and to review performance under the program by providing a critical and frank consideration of two key questions: (i) were the macroeconomic strategy, program design, and financing appropriate to address the challenges faced by the member in line with Fund policy, including exceptional access policy; and (ii) did outcomes under the program meet program objectives? As part of question (i), EPEs were expected to assess the justification for exceptional access, especially

¹ Staff guidance on conducting EPAs was set out in *Operational Guidance for Assessments of Countries with Longer-Term Program Engagement*, August 20, 2003 which was subsequently revised in *Ex Post Assessments of Members with a Longer-Term Program Engagement—Revised Guidance Note*, February 25, 2010. The latter took account of definitional and procedural changes on LTPE adopted by the Board as well as decisions on easing work pressures.

² Public Information Notice (PIN) No. 03/49, *IMF Concludes Discussion on Prolonged Use of Fund Resources*, April 9, 2003.

³ The requirement for EPEs was agreed by the Board in September 2002 for members using exceptional access in capital account crises, and extended to any use of exceptional access in February 2003. Staff guidance on conducting EPEs was set out in *Ex Post Evaluations of Exceptional Access Arrangements Guidance Note*, August 8, 2005, and *Ex Post Evaluations of Exceptional Access Arrangements—Revised Guidance Note*, issued on February 25, 2010.

with regard to four criteria (exceptional balance of payment pressures, debt sustainability assessment, reaccessing capital markets, and prospects for success in view of shortcomings or risks).

4. Procedural requirements for EPAs—which were also adopted for EPEs—were designed to encourage a fresh perspective and credibility. Reports are prepared by an inter-departmental staff team that should include representatives from the area department as well as one each from the Strategy, Policy and Review Department (SPR) (previously PDR) and at least one other functional department.⁴ Initially, there was only a “best practice” call for the team leader not to be the regular mission chief but this was strengthened in the 2010 revised guidelines to require that the team leader be from outside the home area department. The draft report is subject to the normal internal review process, but the revised guidelines state explicitly that final decisions are the responsibility of the team leader. Surprisingly, however, the draft reports forwarded to management are subject to the same “dual signature” approval—by the head of the home area department (or representative) and the SPR review officer—as all other country reports. This aspect seems inconsistent with the independence of team authorship. Draft reports are also to be discussed with the authorities of the country concerned, but the analysis and conclusions are not to be subject to negotiation.⁵

5. A streamlined EPA Update was introduced in 2009 as part of a broader initiative to reduce staff work pressures.⁶ For countries that continued to meet the LTPE definition after an interval of at least five years, new assessments were required but in many cases these could be shorter Updates.⁷ They are expected to focus on whether staff drew on lessons identified in the first EPA and on macroeconomic performance since then (including results under a Fund-supported program), as well as priority areas for any future program.⁸

6. In terms of coverage, the two types of self-assessment have covered a little over one-third of all agreed programs, a share which has not varied much over time (Table 1). The

⁴ Estimating the resource costs of the self-assessment exercises is beyond the scope of this review, but discussions with various team leaders suggest that a typical team spent about 9 months of staff time preparing the report. The departments providing the individual staff members bear these costs and the home area department bears the associated travel costs. Typically, the latter provided for one short staff visit to the country by the evaluation team leader, but not by the rest of the team.

⁵ More recently, evaluation teams can visit the country for preliminary discussions with authorities (and other stakeholders) prior to drafting the report, but this is not mandatory.

⁶ Updates are expected to be only half the length of a full EPA, i.e. around 2500 words.

⁷ If a program had been cancelled or interrupted for more than six months during the five-year qualifying period, then a full EPA was required.

⁸ In light of the crisis-related demands on staff resources, in 2009 the Board also approved a one-year suspension, until end-August 2010, on the requirement for an EPA update for members with LTPE if the original report was issued more than five years earlier.

share of total IMF lending covered by the assessment exercises is much larger—about three-quarters in recent years—reflecting the fact that all large (“exceptional access”) lending cases are reviewed under the EPE. Almost half of all assessments have involved countries in Africa, but over three quarters of the IMF financing commitments in assessed programs have been to countries in Europe.

Table 1. Self-Assessment of IMF Programs: Coverage, 2003–12
(In percent)

	Full period 2003–12	2003–07	2008–12
Programs covered by EPA or EPE assessments ¹ :			
As a share of number of agreed programs (excluding FCL)	36.1	35.4	36.8
As a share of SDR amount of agreed programs (excluding FCL)	73.5	68.6	74.6

¹ Includes Joint EPA/EPEs and EPA Updates.

II. METHODOLOGY OF THIS REVIEW

7. The principal source of evidence is a desk review of the 49 country self-evaluations undertaken between August 2005 and December 2013, including 3 Joint EPA/EPE reports and 7 EPA Updates (Table 2).⁹ Thirty of these reports cover programs primarily under concessional facilities (the Poverty Reduction and Growth Facility and the Extended Credit Facility) and 19 cover programs under the General Resources Account (GRA). Each self-evaluation report is rated on the quality of the assessment according to a series of criteria covering the following broad headings: (1) the rationale for IMF program involvement; (2) program design; (3) effectiveness of IMF involvement, including the appropriateness of conditionality; (4) forward-looking strategy; and (5) overall judgment on the report, including the extent to which it identifies clear lessons (see Annex 1 for the full set of assessment criteria). A four-point scale (not discussed, weak, adequate, and very good) is used to rate how well the reports performed under each criteria. In order to maximize the scope for comparisons, the criteria have been designed so that most are applicable to both EPAs and EPEs, but some are applicable to only one type of report, reflecting mandates in their respective guidelines. For example, only EPA reports are assessed on how well they investigate the reasons for LTPE and the impact of such engagement on domestic institutions; only EPEs are assessed on how well they investigate the justification for exceptional access and the role of the IMF in managing the crisis. The extent of any

⁹ This cut-off date was chosen in part to cover all EPAs issued since an earlier IMF staff review of EPAs. See *Review of Ex Post Assessments and Issues Relating to the Policy on Longer-Term Program Engagement*, March 20, 2006. It also allows for coverage of EPEs completed since the issuance of the August 2005 Guidance Note on their preparation.

disagreements with the authorities or within the IMF Board (based on a review of the contents of the Summing Up) is also rated. This desk review cannot assess whether particular judgments made in the reports are correct—since this would require a separate country evaluation—but focuses on how well particular issues are explored, the quality of the supporting analysis, and the clarity of the lessons drawn.

8. Structured interviews were also conducted with senior staff in area and functional departments involved with the EPA and EPE exercises as well as with a number of evaluation team leaders. The interviews were based on a standard set of questions sent in advance that covered (i) the usefulness of the EPA and EPE exercises as tools for internal self-evaluation and learning as well as areas for improvement; (ii) appropriateness of the country coverage; (iii) independence of the evaluation teams and any constraints on their robustness; and (iv) arrangements for following up on lessons generated by the evaluations.

Table 2. EPAs and EPEs Covered by this Review¹

EPAs (23)	EPA Updates (7)	EPEs (16)	Joint EPAs/EPEs (3)
<i>Burkina Faso (2006)</i> <i>Burundi (2011)</i> <i>Ghana (2007)</i> <i>Guyana (2006)</i> <i>Jordan (2005)</i> <i>Kenya (2008)</i> <i>Kyrgyz Rep. (2011)</i> <i>Lao PDR (2006)</i> <i>Malawi (2012)</i> <i>Mali (2011)</i> <i>Mauritania (2011)</i> <i>Mongolia (2005)</i> <i>Nicaragua (2007, 2012)</i> <i>Niger (2011)</i> <i>Pakistan (2005)</i> <i>Rwanda (2006)</i> <i>Sao Tome and Principe (2012)</i> <i>Senegal (2006)</i> <i>Tajikistan (2006)</i> <i>Tanzania (2006)</i> <i>Togo (2006)</i> <i>Ukraine (2005)</i>	<i>Armenia (2012)</i> <i>Benin(2010)</i> <i>Burkina Faso (2013)</i> <i>Gambia (2010)</i> <i>Georgia (2011)</i> <i>Sierra Leone 2010)</i> <i>Zambia (2011)</i>	<i>Belarus (2010)</i> <i>Brazil (2006)</i> <i>Costa Rica (2011)</i> <i>El Salvador (2010)</i> <i>Greece (2013)</i> <i>Guatemala (2011)</i> <i>Hungary (2011)</i> <i>Latvia (2012)</i> <i>Romania (2012)</i> <i>Sri Lanka (2013)</i> <i>Mongolia (2011)</i> <i>Iceland (2012)</i> <i>Pakistan (2012)</i> <i>Ukraine (2011, 2013)</i> <i>Uruguay (2007)</i>	<i>Argentina (2006)</i> <i>Serbia (2011)</i> <i>Turkey (2008)</i>

¹ The review covers EPAs, EPA Updates, EPEs, and Joint EPA/EPEs issued between August 1, 2005 and December 10, 2013.

III. EFFECTIVENESS OF THE SELF-EVALUATIONS

A. How Well Do the Self-Evaluations Cover Different Aspects of the Program Engagement?

9. The evaluation reports were rated for how well they assessed different aspects of the IMF engagement in the member country, including the rationale for IMF program involvement, program design, effectiveness of IMF involvement, and discussions of the forward-looking strategy. This section discusses the overall results of the desk reviews and draws on specific examples as well as views of interviewees to illustrate particular strengths and weaknesses that underlie the ratings.

10. **The reports are better at assessing details of the program design and its implementation than at questioning fundamental assumptions underlying the overall strategy and considering possible alternative approaches.** Most reports, especially the EPEs, do at least an adequate job of assessing whether IMF program engagement was justified, but most EPAs do not question the underlying incentives leading to prolonged program engagement and largely take for granted such engagement. One-quarter of all EPAs, but no EPEs, are rated as having a weak assessment of program involvement (Figure 1). The reports rated as weak typically do not question continued program involvement in spite of limited progress, especially on structural reforms, over a long stretch of previous programs. However, even among reports with an adequate or better discussion of the issue, those covering low-income countries mostly accept without serious questioning that the Fund has a role in helping these countries address continuing structural and institutional challenges as well as in providing signals to donors on the adequacy of the policy framework and that this role is best met through a lending arrangement. Only a few raised questions about the broader systemic incentives that led to LTPE. Examples of these few included the Guyana EPA, which frankly discussed the role played by HIPC debt relief: “... *with debt relief at stake, both sides were under strong pressure to continue a program relationship that lacked both realism and ownership.*” Among GRA cases, the Argentina report concluded that the main reason for programs after 2002—in the face of wide disagreements on policies between the authorities and the IMF—was to avoid Argentina defaulting on its Fund obligations. A number of interviewees suggested that, while the guidance notes for conducting EPAs and EPEs give no indication of whether such broader questions should be addressed, de facto incentives are usually to keep the focus of the evaluations fairly narrow, since most junior team leaders would feel constrained in addressing broader controversial issues.

11. **Almost all reports assess adequately or very well the macroeconomic challenges and strategy as well as the appropriateness of macro conditionality in programs** (Figure 1). The one area of macro conditionality that is not always assessed well concerns

ceilings on the government wage bill. Although this is an area of some controversy,¹⁰ about half of the reports on countries with some form of wage bill conditionality merely describe it, without any substantive assessment. An example of a report that did investigate the issue is the Benin EPA, which concluded that a wage bill ceiling was not an effective instrument when broader civil service reform was needed.

12. EPAs generally assess poorly the appropriateness of the targeted mix between adjustment and financing whereas EPEs handle the issue adequately or very well.

Two-thirds of EPAs discuss poorly the issue—with no substantive consideration of alternative adjustment paths—and none (with the notable exception of the joint EPA/EPE reports) are rated as giving a very good assessment (Figure 1). Even when the adjustment/financing mix is discussed in EPAs, it is typically couched in terms of potential debt distress implying the need for greater emphasis on concessional financing. Only rarely are questions raised about the tradeoffs between the targeted fiscal path, growth, and external financing strategies. For example, few reports discuss the tradeoffs between borrowing and the level of productive spending, even though in some cases the authorities explicitly raised this issue in their comments on the draft evaluation (e.g., Burundi, Guyana). However, a number of reports (e.g., the Armenia EPA Update) do make use of cross-country comparisons of the extent of targeted fiscal adjustment. In contrast, most EPEs discuss the adjustment/financing mix at least adequately—which is perhaps not surprising given the centrality of the issue to the justification for exceptional access. For example, the Uruguay and Guatemala EPEs have an explicit analysis of whether the size of targeted fiscal adjustment was appropriate.

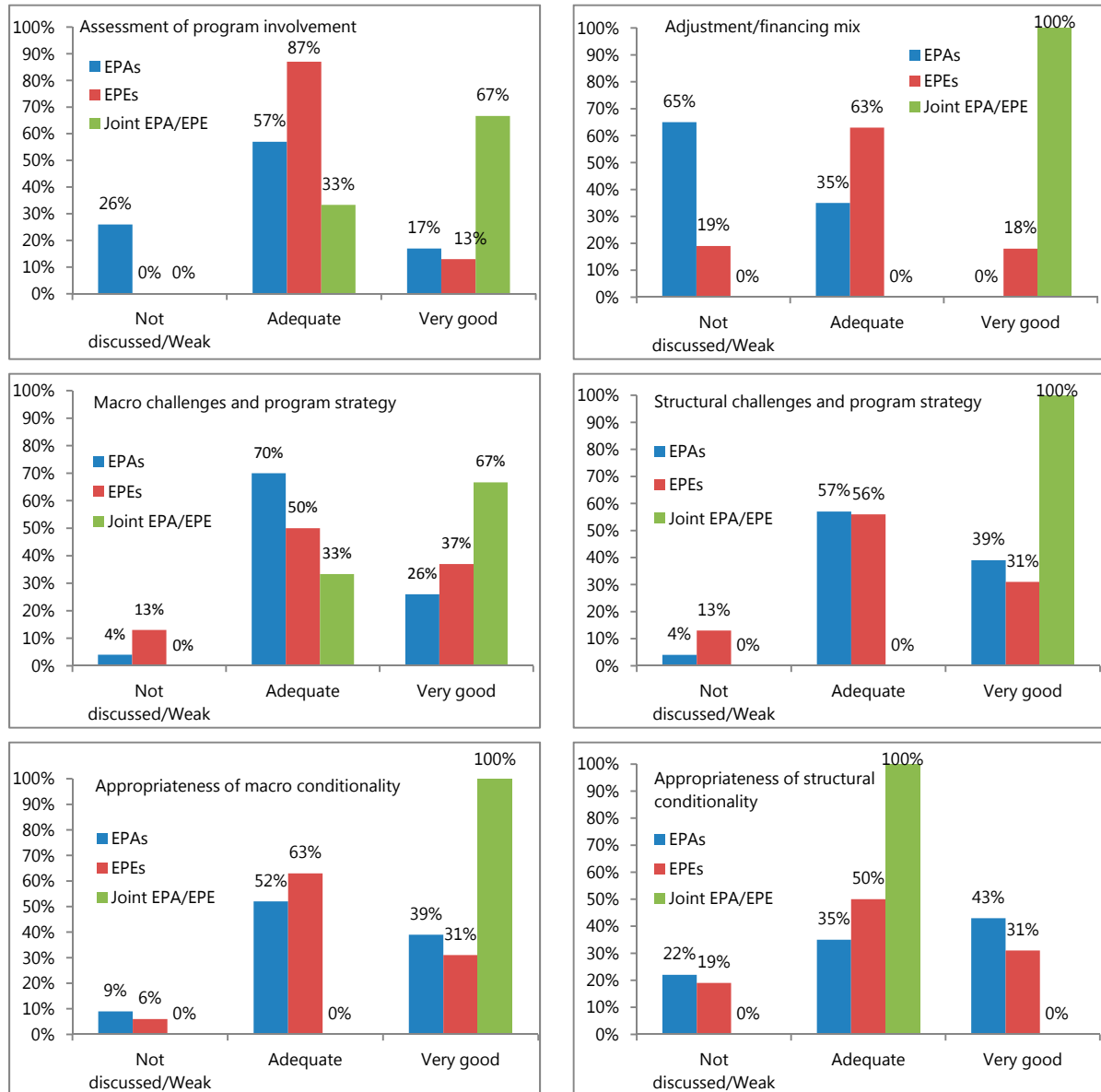
13. Most evaluations discuss at least adequately structural issues and the appropriateness of structural conditionality, but about one-fifth of reports handle these issues poorly. This was one of the areas with the greatest variability in the quality of assessments—amongst both EPAs and EPEs. Some reports merely described the structural issue and associated conditionality, with no questioning of design even with the benefit of hindsight. In contrast, among the approximately one-third of all reports rated as discussing the appropriateness of structural conditionality very well were some that contained quite detailed analysis (drawing, for example, on cross-country data on the extent of structural conditionality) and pointed conclusions about the approach taken by the Fund (Figure 1).¹¹ For example, the Kenya EPA concluded that the Fund’s approach to major incidents of corruption (i.e., insisting on prosecution of former high-level officials) was not macro-critical and delayed the focus on institutional strengthening. The Pakistan EPA discusses in depth the

¹⁰ See, for example, *Does The IMF Constrain Health Spending in Poor Countries?*, Center for Global Development, June 2007.

¹¹ Thirty-nine percent of EPA reports and 31 percent of EPEs are rated as discussing the appropriateness of structural conditionality very well.

pros and cons of extensive structural conditionality in early programs and, emphasizing the difficulty of finding macro-critical individual conditions even when the structural issue itself (e.g., tax reform) is critical, concludes that greater parsimony would have been justified. The Guatemala 2011 EPE discusses in some depth the rationale of not targeting more radical structural reforms beyond the immediate crisis-oriented response as well as the pros and cons of structural conditionality targeting tax reform.

Figure 1. Content of Discussion of Program Involvement and Program Design
(In percent of total reports, excluding EPA Updates)

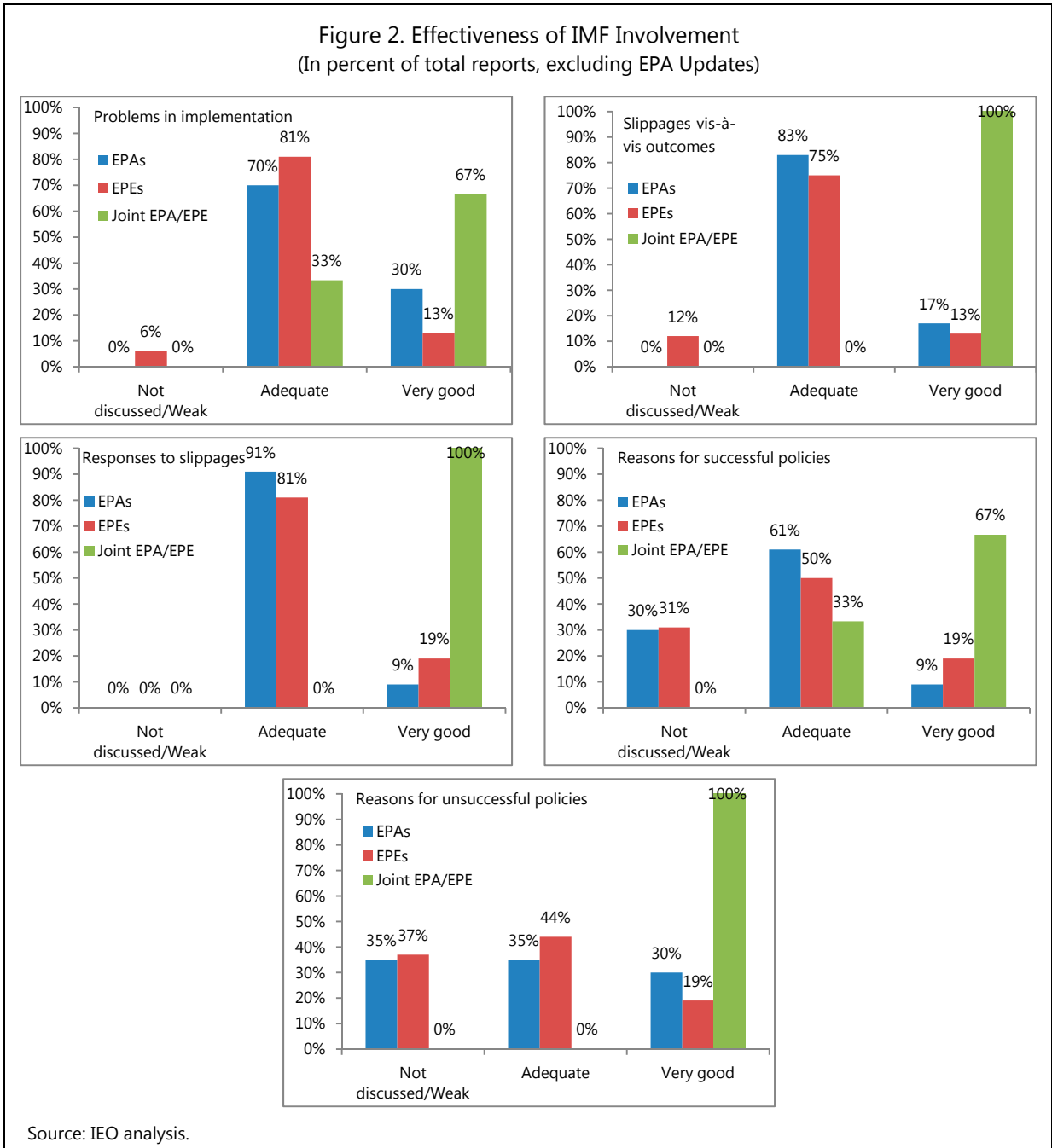


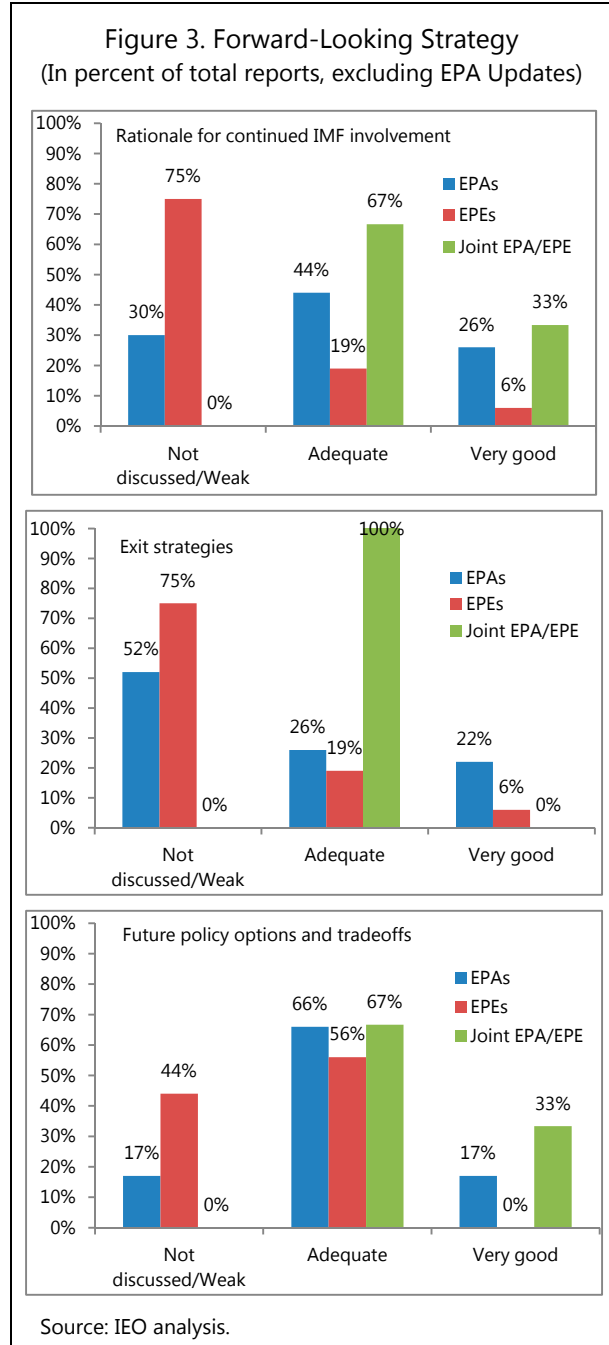
Source: IEO analysis.

14. **Almost all the evaluations document adequately or very well what happened during program implementation. However, the reports' track record in exploring the reasons for successful and unsuccessful policies is more mixed.** Almost all reports are rated as having at least an adequate discussion of problems in program implementation as well as of slippages vis-à-vis targeted outcomes and responses to those slippages (Figure 2). Indeed, a number of reports, especially among the EPAs, are overly detailed in their program accounting. In other words, the reports answer the “what happened?” question quite well. But the extent to which they explore the “why?” and “what if” questions is more varied; for example, about one-third of EPA and EPE reports do not discuss adequately the reasons for unsuccessful policies (Figure 2). Lack of ownership is the most frequent reason given for implementation problems, but many reports do not go beyond this ownership mantra to probe what underlies the stated lack of ownership and what it implies for the program design and the tradeoffs facing the IMF. Among those that do probe further, the usual conclusion is that a greater use of prior actions, or a more parsimonious set of reforms, would have been justified. Few reports consider whether a different set of policies (i.e., a broadening of the policy space) could have attracted more ownership or whether the IMF should have refrained from supporting a program. One report that did discuss frankly these tradeoffs was the Guyana EPA, which concluded that the 1998–2001 programs under the PRGF were overly ambitious regarding the pace of reform but also excessively compliant in the face of major slippages. The evaluation attributed this “*problematic mix of ambition and lenience*” to the pressures created by HIPC debt relief and concluded that the IMF should have either accepted a weaker, but more realistically owned, program or disengaged from program involvement. Similarly, the Togo EPA concluded that the IMF should have exited the program relationship sooner, given weak ownership. Sometimes the analysis of the authorities points to possible answers not explored in the assessment report itself. For example, in their comments on the 2012 EPE, the Latvian authorities discuss the question of “why success?” explicitly, concluding that the fact that it was their own program from beginning to end, despite some disagreements with the Fund, the front-loaded nature of the adjustment strategy, and the heavy involvement of regional partners were all important factors contributing to the outcome.

15. **The evaluation reports are better at setting out a future policy agenda for the country than at assessing the future role of the IMF.** The majority of all reports discuss adequately future policy options and tradeoffs whereas only about one-fifth of EPAs and two fifths of EPEs have a weak analysis of such issues (Figure 3). But a majority of reports address poorly the issue of an exit strategy from Fund-supported programs. The guidelines for EPEs do not call for a discussion of the rationale for continued IMF program involvement, so it is not surprising that many of these evaluations do not address the issue well. In contrast, an assessment of continued program engagement is supposed to be at the core of the EPAs, but the reports assess the issue with widely varying effectiveness; 30 percent discuss the issue poorly—essentially assuming a continuation of the Fund’s program role (i.e., a generic “one more program” recommendation) with little questioning of

the tradeoffs involved—and about one quarter discuss the matter very well. The latter often drew on their assessment of previous program performance to identify specific pre-conditions needed for a reasonable assurance of success in subsequent programs.





16. **Almost half of the EPAs and one-third of the EPEs reviewed fail to identify clear lessons (Figure 4).** There is also considerable variation in the reports’ effectiveness in addressing different types of lessons—an issue that will be discussed below. The weaker reports either identify no lessons or ones that are so general as to be not helpful for future IMF operations. Some other reports had insightful discussions of specific issues, but then pulled their punches when identifying concrete lessons. For example, the Turkey EPA/EPE has a high-quality analysis of many fundamental issues, including the extent of structural conditionality (i.e., the so-called “Turkey exception” of detailed, roadmap conditionality),

but does not go the last mile in terms of drawing pointed lessons for the IMF on whether this approach was appropriate. In contrast, about one quarter to one-third of reports do identify clear and actionable lessons. For example, the Brazil EPE concludes that “light” program conditionality was justified at the initial stage but that more could have been done at the time of the program extension to address longer-term problems. The Hungary EPE concludes the significant underestimation of the broader effects of the global de-leveraging on Hungary, which led to over-optimism about short-term growth prospects, illustrates the need for a greater focus on macro-financial linkages in programs.

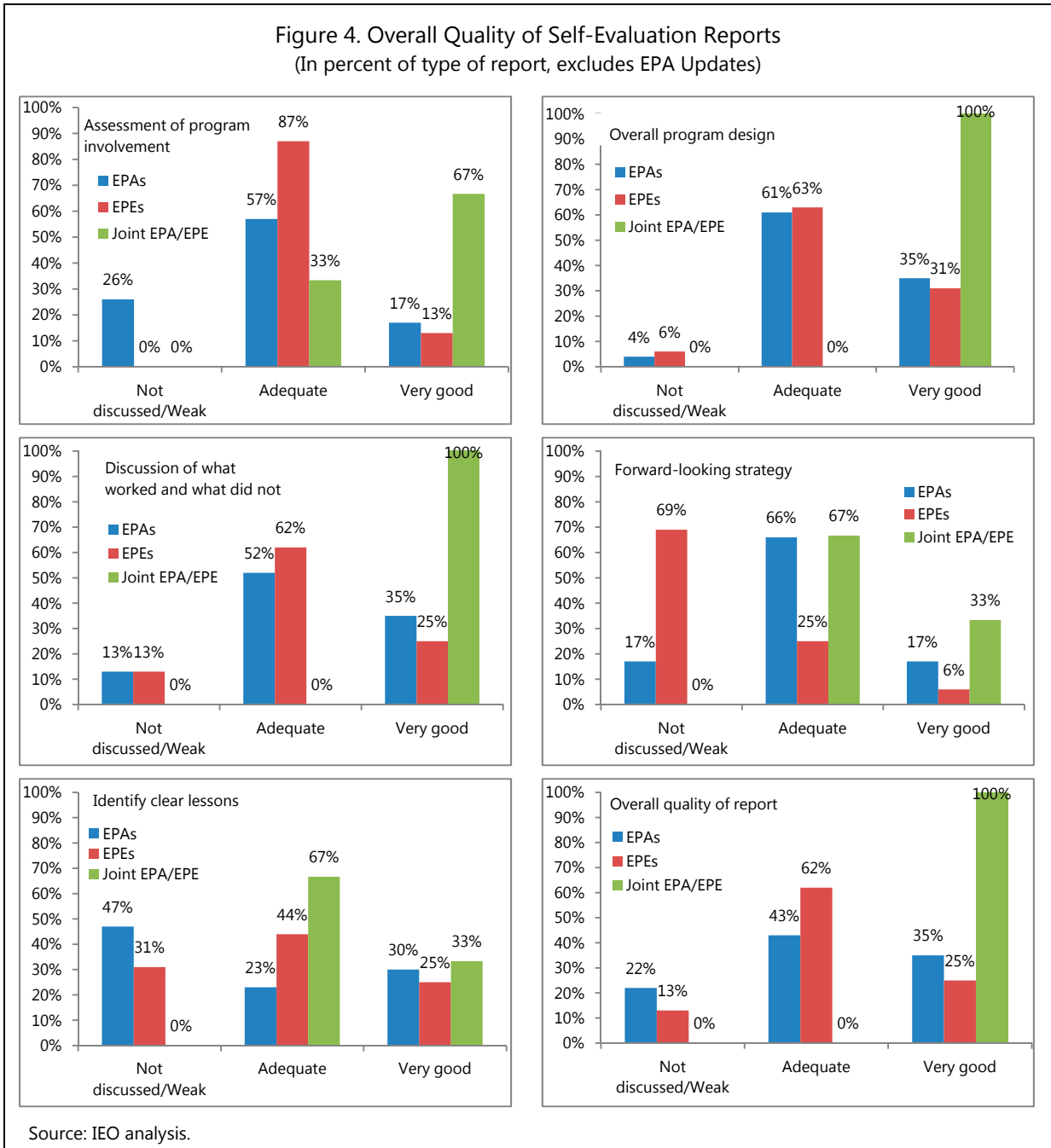
17. **Few reports integrate the authorities’ views into the overall assessment.** Of the 49 reports, 34 are rated as having no or a weak discussion of the authorities’ views; only two do a very good job by considering these views as part of the comprehensive assessment. If the authorities’ views are given at all, they are usually contained in an annex, with limited references in the main report.¹² Even reports that are otherwise strong analytically often do not consider the authorities’ specific criticisms of program design and say whether these criticisms are valid or not. For example, in their response to the Argentina EPA/EPE, the authorities made a very long list of objections—including to the assessments of monetary policy, structural reform, and the approach to lending into arrears—which were not discussed in the main text. Similarly, the comments by the Latvian authorities in the annex to the 2012 EPE assert that the program’s mis-estimation of the size of output collapse and subsequent export response suggested that “*the staff had misread how economies under fixed exchange rates adjust.*” It would have been helpful from a learning perspective to know what the self-assessment teams thought of these criticisms. In a few instances, the authorities made specific suggestions (e.g., the need for greater staff continuity in their response to the Nicaragua EPA) that were not picked up in the main report.

18. **The views of stakeholders other than the authorities and the IMF staff are largely absent from the evaluations.** Very few reports discuss the views of the World Bank or other donors and hardly any mention the views of other local stakeholders. This was sometimes the case even when the annex said that other stakeholders’ views had been sought (e.g., the Sri Lanka EPE). (The Kenya EPA is an exception.)

19. **In sum, there is wide variation in the overall quality of the evaluation reports, especially among EPAs.** While the large majority of reports were rated of at least adequate quality, about one-fifth of EPAs and a small number of EPEs were judged to be weak (Figure 4). These latter reports typically just describe what happened under the various programs, with little real assessment of the Fund’s role and few identifiable lessons. Implementation problems are usually attributed to lack of government ownership, but with

¹² This largely reflects the timing and budgetary constraints on the assessment process. As described in Section I, the typical procedure is for the staff team to prepare the draft report at headquarters, with little initial input from the authorities, and for the team head then to visit the country for a few days to discuss the report.

little exploration of what this might imply for the original program design. In contrast, about one-third of reports (and all three of the joint EPA/EPE reports) were judged to be of very good quality.



20. **The two factors most strongly associated with the overall quality of the evaluations and the clarity of the lessons drawn appear to be the level of experience of the team leaders and the type of lending arrangements being reviewed.** Reports prepared by team leaders in the B grades are rated substantially higher for the overall quality of the

evaluation and the clarity of the lessons identified compared with reports prepared by more junior (A14-A15) team leaders (Figure 5).¹³ This may be because more senior staff have the necessary experience to identify problems or because they are less inhibited in flagging such problems. It could also be because more senior staff are assigned to evaluate cases that are known in advance to be “difficult” and hence more likely to generate substantive criticisms and lessons. Evaluation reports prepared by teams led by the regular mission chief—a now discontinued practice—were not markedly weaker in overall quality than those prepared by a team leader from outside the department.¹⁴ The particular outside department (e.g., SPR or other functional department) did not appear to make any systematic difference to the ratings. Surprisingly, evaluations led from within the area department but not by the mission chief, produced the strongest reports of all, but the sample is small (seven reports). The clarity of lessons and overall quality of the report are both somewhat higher for GRA cases than for programs supported under the concessional (PRGF/ECF) facilities; for example, one-half of the latter but only one-quarter of the former had weak lessons.

21. Despite a few candid and forceful EPEs reviewing recent country programs in response to the global crisis, the evidence does not suggest any marked shift in the average quality of the self-assessments in recent years—with the exception of some moderate improvement in the EPA Updates. Testing for possible shifts in the quality of the self-assessment reports associated with various developments in recent years—such as the adoption of revised guidelines on EPAs (February 2010) or the shift in coverage of programs assessed under the EPE to those associated with the global crisis—did not indicate any overall gain in clarity of lessons or quality of the reports. However, several recent reports—notably the Greece 2013 EPE and the Ukraine 2013 EPE—did contain very frank analysis and pointed lessons.¹⁵

B. EPE- and EPA-Specific Issues

22. EPE reports handle adequately or very well evaluation issues specific to the exceptional access cases. These issues include an assessment of the root causes of the vulnerability, coherence of policies to address these vulnerabilities, and factors influencing capital account outturns (Figure 6). About half of all the EPE reports assessed very well the coherence of policies/conditionalities in addressing the underlying weaknesses that contributed to crisis vulnerability. A number of reports are especially effective at assessing

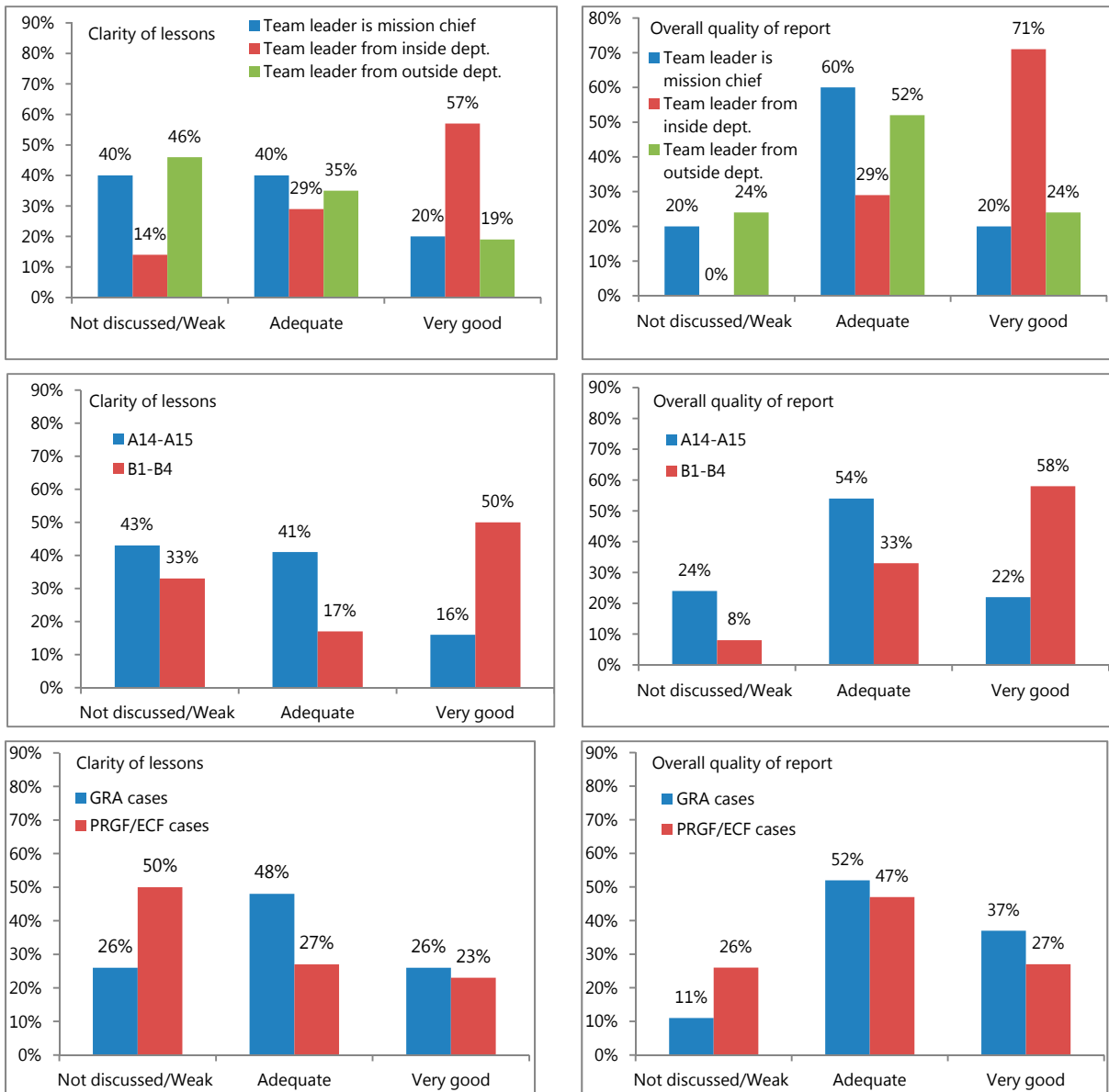
¹³ Of the 49 evaluations reviewed, one was headed at the B4 level, 2 at the B3 level; 8 at the B2 level; one at the B1 level; 36 at the A15 level; and 2 at the A14 level.

¹⁴ Five evaluations were led by the regular mission chief (all in the earlier part of the period covered); 7 were led from within the same area department but not by the mission chief; and 37 had team leaders from outside the area department. About half of the latter were from SPR.

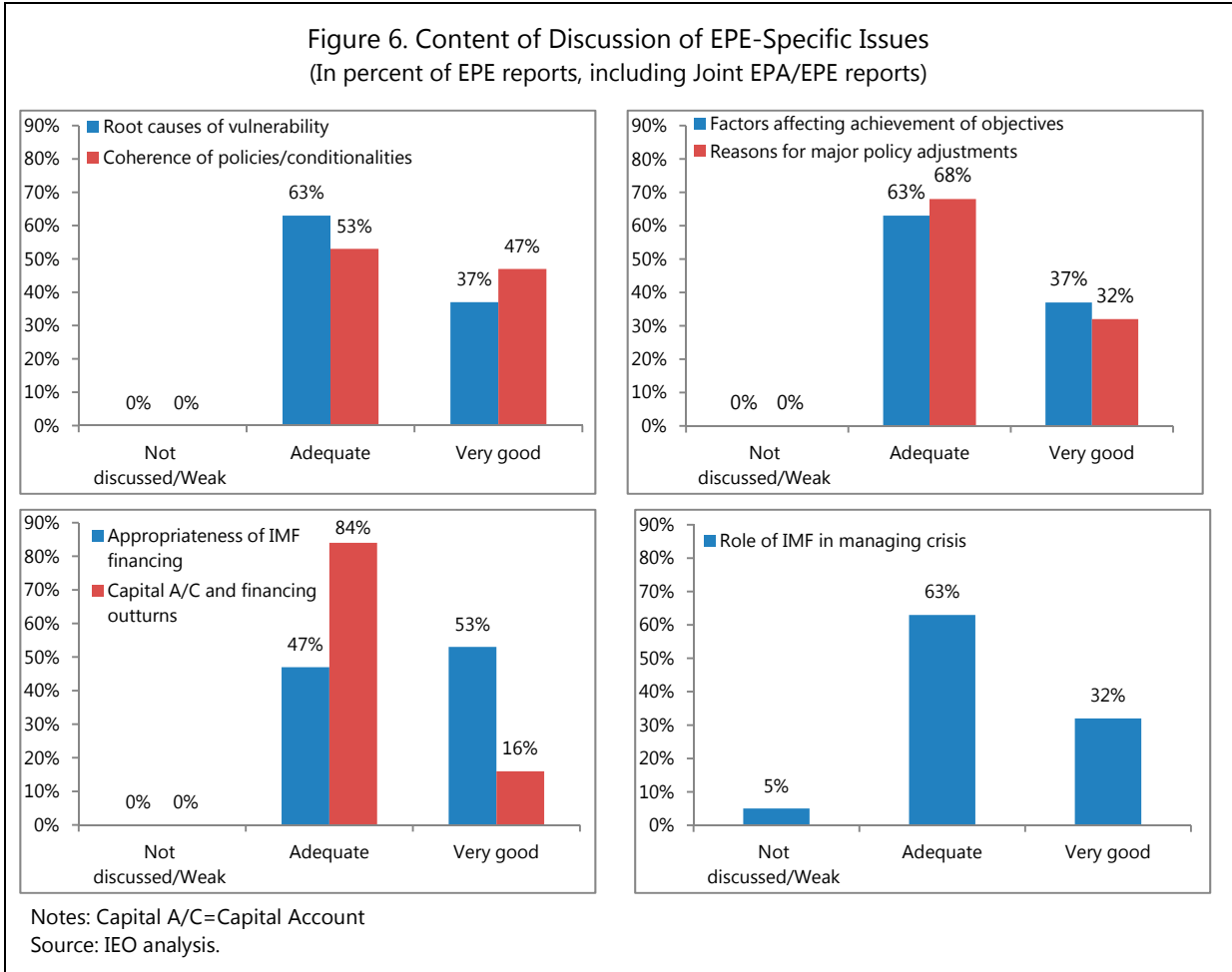
¹⁵ As noted earlier, it is beyond the scope of this paper to evaluate the merits of the specific conclusions reached in these papers.

the appropriateness of the size and phasing of IMF financing, including the justification for exceptional access vis-à-vis the criteria set out in the guidelines. For example, the Brazil EPE discusses the size of financing in relation to potential pressures on the capital account and concludes that the IMF was correct to choose a size of financing that exceeded market expectations, reflecting a judgment that the crisis was essentially one of confidence, but points to the risks involved in such an approach—risks that it concludes the IMF was justified in taking.

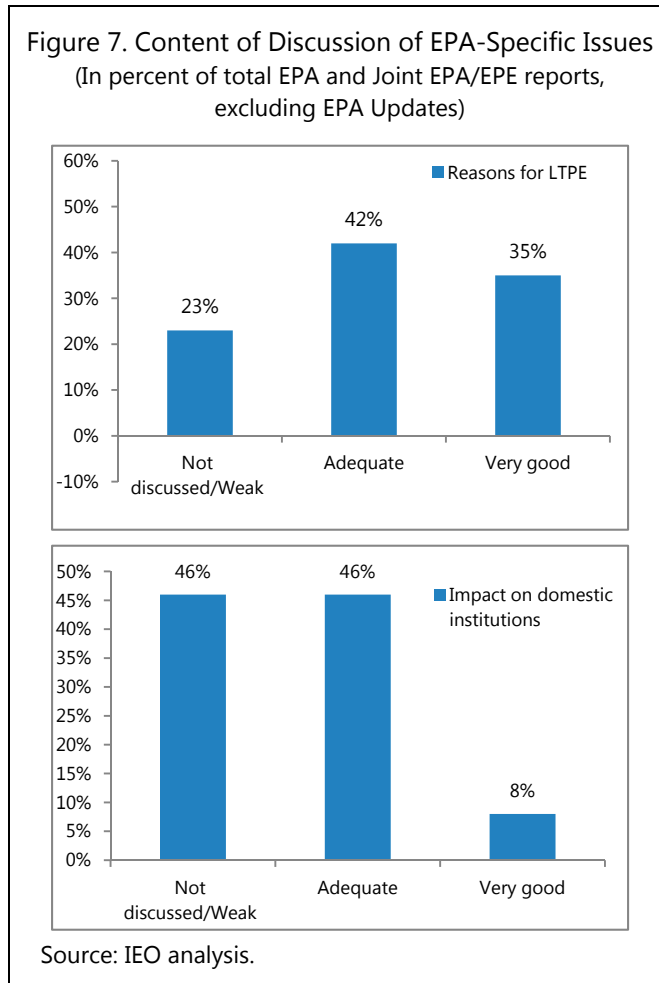
Figure 5. Clarity of Lessons and Quality of Overall Report by Type of Team Leader and Type of Country (In percent of reports)



Source: IEO analysis.

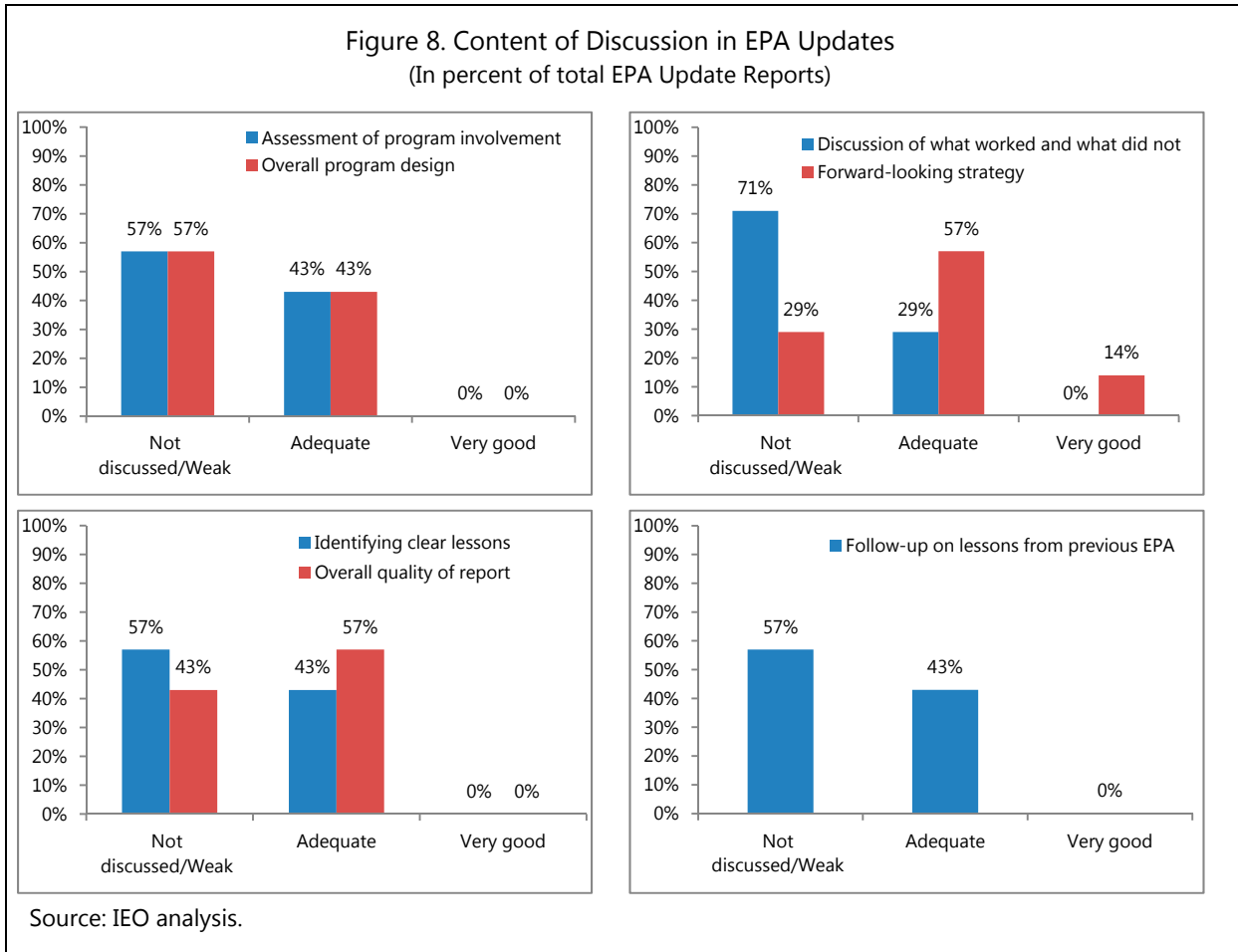


23. **The track record of EPAs in assessing the reasons for longer-term program engagement is mixed, and about half of all reports are weak at assessing the impact of such engagement on domestic policy-making institutions (Figure 7).** Those evaluations that do address the impact of LTPE on domestic institutions usually restrict themselves to discussing the contribution of IMF technical assistance; few make any attempt to assess whether prolonged program engagement has adversely affected domestic policy-making processes. One report even stated that the IMF had provided the policy framework for over two decades, so there was limited domestic capacity to substitute, concluding that this justified a further program. The Ukraine EPA was one of the few to address some aspects of the issue substantively, concluding that repeated programs have weakened the IMF negotiating stance as well as the effectiveness of a program as a commitment device.



24. **EPA Updates are significantly weaker as an assessment tool than other types of evaluation reports.** Although based on a small sample (seven reports), the Updates scored poorly against most criteria, including assessment of program involvement, program design, and identifying clear lessons—with over half of the updates rated weak in these categories (Figure 8). While this is perhaps not surprising given the more streamlined focus (and length) of these evaluations, **they were also not especially effective in following up on whether IMF staff drew on lessons identified in the first EPA—which was an explicit part of their mandate over half had a weak discussion of these issues.** For example, the Gambia report has a “box” summarizing the lessons from the 2005 EPA, but these are not referred to much in the rest of the report; the discussion of fiscal structural reforms does not give a clear assessment of whether application of the 2005 “lessons” delivered better results. In some cases, the lack of systematic follow-up reflects the rather general nature of the earlier EPA lessons. For instance, the Sierra Leone 2010 and Georgia 2011 Updates include a table tracking what has happened vis-à-vis key recommendations of the earlier EPAs, but many of these earlier lessons are very general and would apply to most low-income countries. However, some of the more recent Updates do incorporate more concrete lessons. For

example, the Burkina Faso 2013 Update includes an analysis of structural conditionality design, including some forward-looking lessons.



C. Quality of Different Types of Lessons

25. An assessment of different types of lessons included in the reports highlights sizable differences in their effectiveness in generating clear, specific and operationally relevant messages of different types (Figure 9).¹⁶

- The self-evaluations are most effective in a “stock-taking” of country performance under programs and in checking accountability vis-à-vis various IMF policy guidelines.** All of the reports do at least an adequate (and more than half do a good) job of assessing what was achieved under programs in terms of macroeconomic and structural adjustment, growth, and (where relevant) poverty

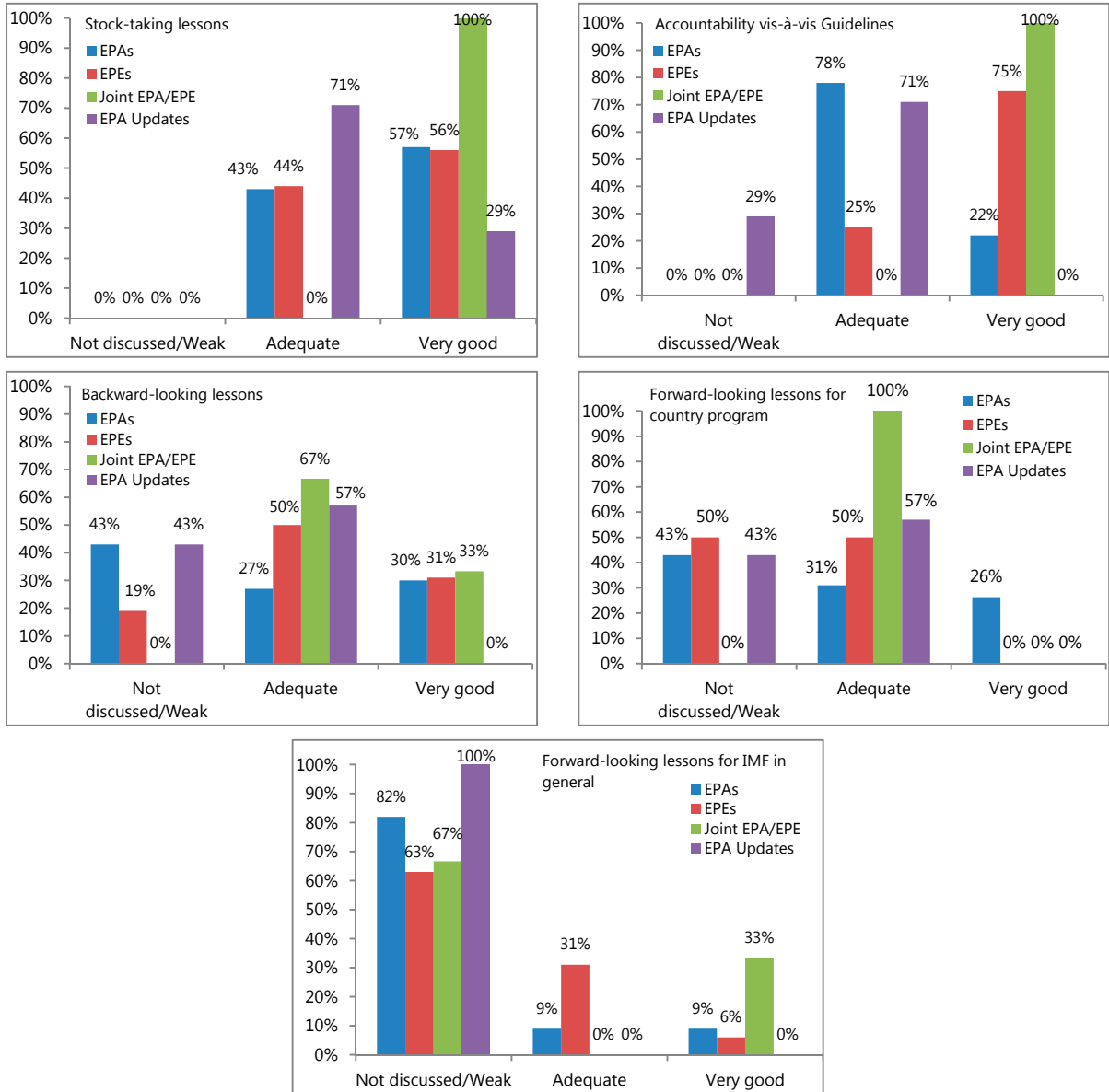
¹⁶ Using the same four-point rating scale as in the previous discussion. A specific type of lesson was rated as discussed very well if the lessons were clear, specific and of relevance to the operations of the Fund.

reduction. Virtually all reports assess adequately or very well accountability against various guidelines. For example, the EPEs are especially effective in assessing the justification for exceptional access, in particular vis-à-vis the four criteria identified in the guidelines.¹⁷

- There is much greater variation in the quality of “backward-looking” lessons. Over two-fifths of EPAs are rated as not generating effective lessons for how the IMF might have acted differently or for identifying specific IMF actions that contributed to a successful outcome. Most often, this reflected a failure to address frankly the “why” and “what if” questions.** In such cases, the facts concerning a performance shortcoming are usually well discussed, but there is limited exploration of how variations in program design might have affected the outcome. For example, many of these evaluations simply state that some key elements of the program were not implemented in a timely fashion because of a lack of ownership on the part of the authorities. While probably true, such statements alone do not add much value to an assessment of the Fund’s role. In some cases, thoughtful discussions in the main report (e.g., in the Ghana report, on the failure of structural conditionality in the form of prior actions on specific price hikes to yield broader price reform) could have been the basis for more pointed lessons. In contrast, about 30 percent of EPAs identify clear lessons on what the IMF might have done differently. For example, the Togo EPA, discussing a case where long-term program engagement largely failed, points to concrete program design changes that could have helped to improve outcomes while making clear that design changes alone would not have overcome fundamental political obstacles and concludes the IMF should have exited the program relationship sooner. The Pakistan EPA concludes that the IMF was wrong to set its own conditionality in the power sector and should have focused more forcefully on the overall quasi-fiscal deficit while asking the World Bank to design specific power sector measures. On average, the EPEs reviewed do a better job of identifying concrete lessons on what worked and what did not, with only one-fifth rated as not discussing such lessons well. For example, the Belarus EPE concludes that the staff initially underestimated the importance of lending under government programs, which should have been directly covered by macro conditionality. The Sri Lanka EPE has a good discussion of why fiscal adjustment was of a lower quality than envisaged.

¹⁷ The four criteria are: exceptional balance of payments pressures on the capital account; debt sustainability assessment; reaccessing capital markets; and prospects for success.

Figure 9. Quality of Different Types of Lessons in Self-Evaluation Reports
(In percent of total number of each type of report)



Source: IEO analysis.

- There was also wide variation in how well the evaluations apply their reviews of performance to forward-looking lessons for the Fund’s future engagement with the member country.** Almost half of the EPAs and EPEs reviewed were rated as not discussing well such lessons. Even in reports that set out a fairly comprehensive agenda for future policies, many do not address in any substantive way—beyond a call for greater ownership—the question of why reforms that have not been implemented in the past should have a different fate in the future. In contrast, about

one-quarter of the EPAs are very effective in linking their assessment of program design and the Fund's role to the forward-looking policy agenda and the terms of future IMF engagement. For example, the Rwanda EPA proposes steps to minimize the risks to implementation of key structural reforms through use of a limited number of prior actions, more closely coordinated with the World Bank. The Guyana EPA has a frank discussion of the tradeoffs facing the IMF and the authorities and the implications of different options for future engagement. On average, EPEs perform slightly worse than EPAs in identifying forward-looking lessons for the Fund's future engagement with the member country, with none rated as discussing such lessons very well. However, this is not surprising because—unlike EPAs—drawing lessons on future engagement is not mentioned as a specific objective in the Guidance Notes for such evaluations.

- **While only a small sample, all three of the joint EPA/EPE reports do at least an adequate job of identifying both backward- and forward-looking lessons for IMF operations in the country concerned.** For example, the Argentina report is brutally frank in discussing the considerations that led the IMF to continue with a program engagement even when it was clear that the authorities did not believe in large parts of the program design.
- **Few self-evaluations identify explicit forward-looking lessons for IMF operations in general.** Of the 49 reports reviewed, only 10 discuss such lessons adequately or very well. Examples include the Pakistan EPA, which has a thoughtful discussion on the complexities of designing effective structural conditionality and on the tensions between short- and long-term reform design and the implications for collaboration with the World Bank; the Kenya EPA, which suggests that multi-year arrangements are not the most appropriate instrument in cases where there are serious governance concerns and a poor track record; and the Hungary EPE, which draws lessons on how to improve EU-IMF collaboration on program design and implementation in crisis cases. The Argentina joint EPA/EPE contains a frank discussion of how the Argentina case shows limitations in the existing IMF policy on Lending into Arrears. Perhaps most notably, the Greece EPE draws a series of potential lessons concerning the need to avoid undue delays in debt restructuring, for more attention to the political economy of adjustment, for more parsimony in fiscal structural reforms, and for streamlining the process of cooperation with European institutions.

D. Views of Interviewees on the EPA and EPE Processes

26. **Almost all interviewees thought the EPAs and EPEs were useful learning exercises, but many said that the process could be improved to add greater value.** This view was reinforced by the survey results conducted as part of the broader evaluation for which this desk review was prepared; of those surveyed who had participated in conducting an EPA or EPE, most rated the exercise as very positive. Most interviewees' comments focused on the

issues of country coverage and timing; choice of evaluation team leaders; and processes for identifying key evaluation questions and incorporating the views of authorities. Interviewees' views on the follow-up to evaluation lessons will be discussed in the next section.

27. **Most of those interviewed said the combined country coverage of the two self-evaluation exercises was broadly appropriate from a risk-based management perspective, but that greater flexibility in conducting the timing of the assessments would improve their value.** The one potential gap in country coverage mentioned by a number of interviewees was mid-sized, non program-intensive cases; such countries tended to be overlooked both by the EPA/EPE process as well as by the triennial reviews of surveillance. In contrast, some thought repeat assessments of LTPE cases risked becoming routine, with limited impact. Many interviewees thought that the existing rules often resulted in assessments that were poorly timed for lesson learning. In particular, the timetable for EPEs in countries with back-to-back programs was thought to be too rushed, resulting in assessments that were shoe-horned into a program schedule that did not allow for lessons to be incorporated into any follow-up program. One possible innovation that might increase lesson learning would be to undertake joint evaluations of a group of countries with similar issues.

28. **The system for choosing evaluation team leaders was widely judged to be inadequate.** Many interviewees thought the usual system—which relied on the senior personnel manager in the home area department informally contacting counterparts in other departments to identify potential candidates—was too ad hoc. The heavy reliance on A15 team leaders who were often leading their first mission was seen by many to result in leaders with too little program experience to examine complex cases. Many would prefer a system based on a roster of potential team leaders judged to have the necessary analytical background and experience.¹⁸

29. **The most frequently mentioned problem with the current evaluation processes was the lack of sufficient prior consultation with the authorities.** Team leaders interviewed said that, in practice, the timetables allowed little scope for incorporating the authorities' views into the report. Some thought that greater prior consultation with the authorities on what were the key evaluation questions would be helpful. The guidance notes left many other questions on this aspect unanswered (e.g., on how to proceed when the government has changed since the period under evaluation.)

30. **Views on the degree of independence of the evaluation exercises were mixed.** Interviewees generally thought there was little direct interference with the assessments of the evaluation teams (although some specific instances of such interference were cited).

¹⁸ However, the survey of IMF staff conducted as part of the self-assessment evaluation indicates that a majority (58 percent) of B-level staff are not interested in participating in a future EPA or EPE exercise, whereas most A-level respondents (56 percent) were interested. This may reflect the view of EPA/EPE team leadership as offering opportunities for career enhancement for senior A-level staff.

Nevertheless, in practice relatively junior team leaders would often be constrained in the extent to which they would criticize the judgments of more senior staff or question broader Fund policies. Many thought that the system of reviewer sign-off on evaluation reports was inconsistent with the stated independence of the teams. Some team leaders said they would have benefited from access to a senior staff member (separate from the country reviewers in the area department and SPR) who would act as a “sponsor” for the integrity of the evaluation process and to whom they could address any sensitive questions about procedures.

IV. SYSTEMS FOR FOLLOWING UP ON LESSONS ARISING FROM SELF-EVALUATIONS

31. An assessment of the actual impact of the self-evaluations on IMF operations is beyond the scope of this review, but two intermediate questions are of interest: (i) what systems are in place within the Fund for following up on lessons identified in the assessments; and (ii) how effective was the Executive Board in supporting key lessons during its discussion of the evaluation reports.

32. **No formal system to monitor or enforce follow-up exists.** Senior staff from the area departments and SPR all said that they do not have a database of lessons and recommendations from the evaluation reports and do not monitor follow-up in any systematic way. In practice, responsibility for implementing any lessons lies with the area department and SPR senior reviewers as well as the mission chief for the country concerned. However, many interviewees said that internal Fund processes often did not facilitate such follow-up: transition periods between mission chiefs are short and managerial processes to enforce follow-up are absent. In many cases, the next Board paper on the country might summarize key messages from the assessment along with any action being taken in response, but this was not required practice. As a result, the shelf life of an evaluation report was often short.

33. **The Summing Up of Board discussions of the reports is almost always bland, muting any pointed messages emerging from the evaluations and often giving little practical direction to the staff for any follow-up.** In the majority of cases, the Board discussion of the evaluation took place as part of the Article IV consultation and the Board’s conclusions are typically conveyed in one short paragraph at the end of the Summing Up/Press Information Notice for the consultation. These paragraphs are generally uninformative and give little guidance to the staff. Moreover, any outside observer would have a hard time concluding from these summaries what learning had taken place from the evaluation exercise. The Summings Up of “stand alone” Board discussions of evaluations are somewhat more detailed, but also tend not to highlight pointed messages, even in cases where the assessment report generated clear lessons.

34. **The main problem seems to lie with the Board discussions themselves.** To investigate further whether the problem lay with non-substantive Board discussions or failure of the Summing Up to reflect adequately clear messages that were underscored during the discussion, a sample of Executive Board minutes were examined. Since it would be

unrealistic to expect a bland evaluation report to generate a pointed Board discussion, the sample covered 13 cases where the underlying evaluation report had identified some clear lessons.¹⁹ Although a small sample, the results suggest that the main problem lies with the Board discussions themselves. Of the cases reviewed, more than half were rated as weak and only three cases had discussions that were rated as very good, with Directors debating and clearly underscoring messages arising from the evaluations. Two of the latter were stand-alone discussions, and one was linked to an Article IV consultation. Many staff interviewees also expressed disappointment with the quality of the Board discussions of the assessments; several evaluation team leaders said they had been surprised not to be asked questions about what they had thought were important messages arising from their reports.

35. **Although the sample of relevant cases is small, the self-evaluations themselves do not appear to follow-up systematically on the messages from earlier evaluations.** The weakness of the EPA Updates in this area has already been noted. But full evaluations also have a mixed record in investigating whether the lessons from earlier evaluations have been absorbed by the Fund. For example, the 2005 Ukraine EPA concluded with a specific checklist for assessing the risks to the effectiveness of any future program with the Ukraine.²⁰ Unfortunately for learning purposes, a subsequent evaluation—the 2011 Ukraine EPE—cited, but did not apply, the checklist and did not assess whether the Fund staff had used the tests when deciding on the new program. In contrast, the 2013 Ukraine EPE did assess how well the reviewed program had attempted to incorporate lessons from the earlier evaluations, especially with regard to the design of structural conditionality.

V. CONCLUSIONS

36. Four broad conclusions can be emphasized. First, this review suggests that most evaluations are reasonably effective at stocktaking what has been achieved under IMF-supported programs as well as at assessing accountability vis-à-vis specific IMF guidelines. Many evaluations also generate lessons on particular aspects of the IMF strategy in the country concerned and propose a future policy agenda for the country, although a minority fail to do this adequately. However, few reports probe more fundamental questions—either about alternative policy strategies or the broader rationale for IMF engagement—even though the occasion of an evaluation is the obvious time for such questions; if not then, when?

37. Second, the evaluations vary widely in quality, especially in generating concrete lessons for improving IMF operations. So there is considerable scope for raising the quality of

¹⁹ EPAs for Guyana, Kenya, Niger, Rwanda, Togo, and Ukraine; EPEs for Belarus, Brazil, Iceland, and Sri Lanka; joint EPE/EPEs for Argentina and Turkey; and the EPA Update for Burkina Faso.

²⁰ The checklist was: (i) is there sufficient political support behind a program? (ii) are the policy-making capacities to implement a program in place?; and (iii) is the program sufficiently focused on addressing the key obstacles to sustained growth?

weaker evaluations. We examined various potential influences on the quality of assessments and lessons—size of country involved, type of economy, geographic region, type of IMF lending facility involved, whether a follow-up program was likely—and none seemed to be a major determining factor; both good and bad assessment reports have been produced across all these categories. In terms of factors directly under the IMF’s control, however, the most critical factor seems to be the experience and independence of the evaluation team leader.

38. Third, the evaluations as presently structured are very inward looking. They do not provide much of an opportunity for broader analysis and debate of potential criticisms of the Fund approach raised by the authorities and contain almost no input from other stakeholders.

39. Finally, despite the variation in quality, the evaluations in aggregate have generated many lessons of potential value to the IMF. But this value will only be realized if the IMF systematically identifies and follows up on the lessons. Managerial processes for such follow-up are weak. A critical issue is the role of the Executive Board, since the present system seems to yield limited returns for the costs involved in preparing Board papers. The Board discussions of the evaluations rarely provide a clear road map for such lessons, which must make follow-up harder. Either the role of the Board should be strengthened to give clear guidance on such follow-up or resources could be saved by converting the evaluation reports into non-Board papers, with the focus on internal lesson-learning by the staff.

40. Although detailed recommendations are beyond the scope of this review, the discussion does suggest a number of areas for improvement:

- The clearance procedures for evaluation reports seem inconsistent with the stated independence of the evaluation team. If the team leader is solely responsible for the final contents of the evaluation, no further sign-off should be needed.
- A roster of potential evaluation team leaders with the necessary analytical background and experience should be developed.
- The evaluations should seek earlier input by the authorities and assess explicitly any substantive criticisms of the IMF approach they raise.
- The role of the EPA Updates needs revisiting—either to make them more effective at the tasks assigned to them, especially follow-up on lessons from earlier EPAs, or to drop the requirement for such frequent assessment of LTPE cases.
- Decide whether the Executive Board should continue to be directly involved in considering the evaluation reports. If they are, a more systematic and pointed discussion on key lessons arising from the reports is needed.
- A formal system for recording key lessons/recommendations from the evaluations and for monitoring follow-up is needed.

ANNEX 1. COUNTRY ASSESSMENT MATRIX FOR REVIEW OF EPAs AND EPES

Topic	Rating	Comments
I. Process Issues		
Team leader ¹		
Organization of Report		
(i) <i>Thematic/chronological</i>		
(ii) <i>Comprehensive/core</i>		
Views of others discussed:		
(i) <i>Authorities</i>		
(ii) <i>World Bank</i>		
(iii) <i>Other stakeholders</i>		
II. Program Involvement²		
(i) Rationale for program involvement		
(ii) Reasons for LTPE (EPA only)		
(iii) Impact on domestic institutions (EPA only)		
III. Program Design		
(i) adjustment/financing mix		
(ii) macro challenges and program strategy		
(iii) structural challenges and program strategy		
(iv) external financing strategy		
(v) risks and adequacy of design to address them		
IV. Effectiveness of IMF Involvement		
(i) problems in implementation		
(ii) appropriateness of macro conditionality		
(iii) appropriateness of structural conditionality		
(iv) slippages vis-à-vis outcomes		
(v) responses to slippages		
(vi) reasons for success		
(vii) reasons for failure		
V. Forward-Looking Strategy²		
(i) rationale for continued IMF involvement		
(ii) exit strategy		
(iii) future policy options and tradeoffs		

Topic	Rating	Comments
VI. Overall Judgment on Report		
(i) assessment of program involvement		
(ii) overall program design		
(iii) discussion of what worked and what did not		
(iv) forward-looking strategy		
(v) identifying clear lessons		
(vi) overall quality of report		
VII. Agreement/Disagreement on Lessons³		
(i) with authorities		
(ii) within IMF Board		
VIII. EPE-Specific Issues		
(i) root causes of vulnerabilities		
(ii) coherence of policies/conditionality in addressing weakness		
(iii) factors affecting achievement of objectives		
(iv) appropriateness of size/phasing of IMF financing ⁴		
(v) capital A/C and financing outturns		
(vi) reasons for major policy adjustments		
(vii) role of IMF in managing crisis		

¹ Options are: 1-mission chief; 2-within area department but not mission chief; 3-outside area department.

² Ratings in each of these sections will be on a four-point scale (0-not discussed; 1-weak; 2-adequate; 3-very good).

³ Options are: 0-No disagreement; 1-limited disagreement; 2-considerable disagreement.

⁴ Including justification for exceptional access, drawing on four criteria set out in guidance note.

ANNEX 2. PROGRAM LENDING SELF-ASSESSMENTS, BY COUNTRY, 2003–14

Ex Post Assessments (EPAs)		
Albania	Guinea-Bissau	Nicaragua
Armenia	Guyana	Niger
Azerbaijan	Honduras	Pakistan
Benin	Jordan	Peru
Bolivia	Kazakhstan	Romania
Bulgaria	Kenya	Rwanda
Burkina Faso	Kyrgyz Republic	São Tomé and Príncipe
Burundi	Lao PDR	Senegal
Cambodia	Lesotho	Sierra Leone
Cameroon	Macedonia, FYR	Tajikistan
Chad	Madagascar	Tanzania
Ethiopia	Malawi	Togo
Gambia, The	Mali	Uganda
Georgia	Mauritania	Ukraine
Ghana	Moldova	Uruguay
Grenada	Mongolia	Vietnam
Guinea	Mozambique	Zambia
Ex Post Evaluations (of Exceptional Access) (EPEs)		
Belarus	Hungary	Pakistan
Brazil	Iceland	Romania
Costa Rica	Ireland	Sri Lanka
El Salvador	Latvia	Ukraine
Greece	Macedonia, FYR	Uruguay
Guatemala	Mongolia	
Joint EPA/EPEs		
Argentina		
Serbia		
Turkey		

Source: IMF "Report on the Incidence of Longer-Term Program Engagement," Table 1. Ex Post Assessments Considered by the Executive Board and Member's LTPE Status, September 2014; IMF Institutional Repository