

---

# IEO Retrospective: Ten Years of Independent Evaluation at the IMF

JOANNE SALOP

This chapter reviews the 18 evaluations issued by the IEO since its inception in 2001. It looks at country coverage (advanced vs. developing economies) and evaluative orientation (investigations vs. studies); it concludes that from an institutional risk management perspective greater attention to advanced economies and investigations is warranted. It also looks at evaluation recommendations, where it finds (1) wide variation across individual evaluations and (2) ambiguity in some about the standing of sub-recommendations relative to headline recommendations; it concludes that greater consistency and clarity are needed, including in explaining the links between evaluation findings and recommendations. It finds considerable scope for the IEO to examine internal IMF governance (from the perspective of “Management and below”), especially with respect to questions about exactly how the IMF and who within the IMF decides what position is taken when institutional policies are either not clearly defined or not fully implemented. In light of several evaluations’ findings of a lack of evenhandedness in IMF advice and/or conditionality, it concludes that future evaluations should do more to document and assess cross-country differences in treatment, as a basis for recommending possible remedies.

## Introduction

This chapter discusses the 18 evaluation reports issued by the IEO between its inauguration in 2001 and 2011.<sup>1</sup> It brings together facts and observations on the reports as a basis for informing conversations within the IEO and between the IEO and stakeholders about evaluation strategies and actions going forward. It draws on inputs and comments on earlier drafts from current and former IEO staff and lead authors of the 18 evaluations. Unless otherwise stated the views expressed are solely those of its author, who it must be said as a matter of disclosure was the lead author of one of the 18 evaluations and a contributor to and/or reviewer of several others.

---

<sup>1</sup>See Part IV of this volume for a full list of evaluation reports.

The exercise was commissioned by the IEO as part of a broader effort to take stock of IEO's first 10 years and hence provide a basis for forward thinking. Another motivating factor was the IEO's interest in preserving and promoting institutional memory among IEO staff and managers, including with respect to challenges and lessons learned in carrying out evaluations. Also, an important milestone in IEO's history was the issuance in 2006 of a report of the external evaluation of the IEO—the “Lissakers Report;” five years on, the IEO thought it timely to assess how well it was following up on that report's recommendations.<sup>2</sup>

The retrospective looks at the coverage, evidence, findings, recommendations, and evolution over time of IEO evaluation reports. Its emphasis is on the 18 reports as a group and on differences across evaluations within the group. It does not and is not meant to provide an in-depth review of individual evaluations, or an evaluation of the IEO. Nor does it consider the impact of IEO evaluations or the effectiveness of the IEO. These issues are being addressed in the context of other IEO initiatives.

The remainder of the chapter is organized as follows. The first section discusses the coverage of IEO evaluations, with respect to both countries and evaluation orientation. The second section discusses the evaluations' evidence—the underlying data and methods used by the IEO. The third section discusses the evaluations' findings and recommendations. The fourth section briefly discusses the IEO's evolution over time. The final section summarizes the retrospective's conclusions.

## IEO Evaluations: Coverage

This section looks at the country coverage and evaluation orientation of IEO evaluations to date. It aims to shed light on where the IEO has focused its attention, as a basis for discussions within IEO and with IEO stakeholders about prioritization and evaluation selection going forward. It also aims to facilitate the work of future IEO evaluation teams by providing them with a typology for associating their evaluations with the appropriate evaluation comparators from the IEO's first 10 years.

For ease of reference, Table 8.1 lists the 18 evaluations issued by IEO to date in two columns, dividing them chronologically into the first half of the review period and the second half. This split conveniently corresponds to before and after the Lissakers Report, though of course there was a gray zone between the two periods. (The *Multilateral Surveillance* evaluation, for example, was launched before, drafted in parallel with, and finalized after the Lissakers Report; its inclusion below in the second half of the review period was validated by its lead author who indicated that the Lissakers Report influenced its final shape.)

---

<sup>2</sup>See “Report of the External Evaluation of the Independent Evaluation Office” (Lissakers Report), March 2006. Available at [www.ieo-imf.org/ieo/files/evaluationofieo/032906.pdf](http://www.ieo-imf.org/ieo/files/evaluationofieo/032906.pdf). The Terms of Reference and Summing Up of the Executive Board discussion of the report may be found in Part IV of this volume.

TABLE 8.1

IEO Evaluations to Date	
First Half Evaluations	Second Half Evaluations
<i>Prolonged Use</i> (2002)	<i>Multilateral Surveillance</i> (2006)
<i>Capital Account Crises</i> (2003)	<i>Aid to Sub-Saharan Africa</i> (2007)
<i>Fiscal Adjustment</i> (2003)	<i>Exchange Rate Policy Advice</i> (2007)
<i>PRSPs and the PRGF</i> (2004)	<i>Structural Conditionality</i> (2007)
<i>Argentina</i> (2004)	<i>Governance</i> (2008)
<i>Technical Assistance</i> (2005)	<i>International Trade Policy</i> (2009)
<i>Capital Account Liberalization</i> (2005)	<i>Interactions with Member Countries</i> (2009)
<i>Jordan</i> (2005)	<i>Financial and Economic Crisis</i> (2011)
<i>Financial Sector Assessment Program (FSAP)</i> (2006)	<i>Research</i> (2011)

Note: Dates are taken from print-version covers, and vary somewhat vis-à-vis the dates the reports were discussed by the IMF Board.

## Country Focus

Table 8.2 summarizes the country focus of the 18 evaluations, grouped according to the IMF's standard country classifications. Seven of the evaluations (or 39 percent) covered the entire membership and 10 (55 percent) exclusively covered emerging economies and/or low-income countries. One focused primarily on advanced economies.<sup>3</sup>

Different ways of looking at the distribution of IEO resources across country groups lead to different conclusions. One view focuses on the distribution across countries/groups of the 11 evaluations that concentrated on specific countries/groups. According to this view, over the 10-year retrospective period the IEO devoted about 10 percent of its evaluation resources to IMF work with advanced economies, and the other 90 percent to IMF work with developing countries, especially emerging economies. The main alternative view focuses on how many of the 18 evaluations dealt with a particular country group, including in the context of all-member evaluations. According to this view, 45 percent of the 18 evaluations covered IMF work with advanced economies, and 94 percent covered work with developing countries—still a major difference in coverage between the two country groups, but a significantly smaller one.

When the 18 evaluations are broken down into the two time periods defined above, the share of all-member reports is seen to have increased sharply (Table 8.2). All-member evaluations increased from only 1 in the first period to 6 in the second period, with a corresponding drop in narrowly targeted attention to IMF work with the emerging economies and low-income countries. In the first period,

<sup>3</sup>*Multilateral Surveillance* is counted as an all-member report, although the big picture does not change materially if it is counted as an evaluation covering the advanced economies. According to its lead author, "The focus of the multilateral surveillance evaluation was on large economies, which for the most part meant advanced economies. China was explicitly included among the large economies, though Brazil and India were not (as least to the same extent). At the same time, [the evaluation] also looked at the feedback from multilateral surveillance to bilateral (i.e., bilateral surveillance as a user of [multilateral surveillance] outputs). In this sense, all members were included in the evaluation."

TABLE 8.2

Distribution of IEO Evaluations by Country Grouping			
IEO Reports Classified by Country Coverage	Entire Period	First Half	Second Half
<b>All member countries:</b> <i>FSAP</i> (2006); <i>Multilateral Surveillance</i> (2006); <i>Exchange Rate Policy Advice</i> (2007); <i>Governance</i> (2008); <i>Interactions with Member Countries</i> (2009); <i>International Trade Policy</i> (2009); <i>Research</i> (2011)	7 (39%)	1 (11%)	6 (67%)
<b>Advanced economies:</b> <i>Financial and Economic Crisis</i> (2011)	1 (6%)	—	1 (11%)
<b>Emerging economies:</b> <i>Capital Account Crises</i> (2003); <i>Argentina</i> (2004); <i>Capital Account Liberalization</i> (2005); <i>Jordan</i> (2005)	4 (22%)	4 (44%)	—
<b>Emerging and/or low-income countries:</b> <i>Prolonged Use</i> (2002); <i>Fiscal Adjustment</i> (2003); <i>PRSPs and the PRGF</i> (2004); <i>Technical Assistance</i> (2005); <i>Aid to Sub-Saharan Africa</i> (2007); <i>Structural Conditionality</i> (2007)	6 (33%)	4 (44%)	2 (22%)
<b>Total</b>	18 (100%)	9 (100%)	9 (100%)

8 of the 9 evaluations focused on IMF work with developing countries, and only one (*FSAP*) was an all-member evaluation. In the second period, only 2 evaluations (*Aid to Sub-Saharan Africa* and *Structural Conditionality*) focused exclusively on IMF work with developing countries.<sup>4</sup>

### Evaluative Orientation

With respect to issues and orientation, the retrospective classified 8 of the 18 evaluations as “investigations” of topics where serious prior concerns had surfaced about IMF performance in the areas of institutional governance, financial crises, and operational policies; it classified the other 10 as evaluation “studies.” It found this distinction between evaluation investigations and evaluation studies a useful tool for probing and comparing differences and similarities across evaluations with respect to evidence, findings, and recommendations; for prioritizing among competing evaluation proposals; and for informing debate. This said, it also recognized that there were areas of overlap between the two groups, with, for example, all IEO evaluations highlighting and investigating problems that their examinations of the evidence happened to unearth and studies starting with hypotheses about Fund performance.

When the 18 evaluations are broken down into the two time periods, the share of investigations is seen to have increased (Table 8.3). The share of investigations rose from 33 percent in the first period to 56 percent in the second period driven by the focus in 2007 on evaluations that focused on compliance with IMF operational policies. Table 8.3 also shows a further breakdown into five sub-categories (three for investigations and two for studies): governance, financial crises, operational policy compliance, soft mandates, and activity management.

<sup>4</sup>For the low-income countries the degree of exclusive attention stayed broadly unchanged between the two periods, with the *PRSPs and the PRGF* evaluation in the first period and *Aid to Sub-Saharan Africa* in the second.

TABLE 8.3

	Number of Evaluations (Percent share)		
	Entire Period	First Half	Second Half
<b>Investigations</b>	8 (45%)	3 (33%)	5 (56%)
Governance: <i>Governance</i> (2008)	1 (6%)	—	1 (11%)
Financial crises: <i>Capital Account Crises</i> (2003); <i>Argentina</i> (2004); <i>Financial and Economic Crisis</i> (2011)	3 (17%)	2 (22%)	1 (11%)
Operational policy compliance: <i>Prolonged Use</i> (2002); <i>Aid to Sub-Saharan Africa</i> (2007); <i>Structural Conditionality</i> (2007); <i>Exchange Rate Policy Advice</i> (2007)	4 (22%)	1 (11%)	3 (33%)
<b>Studies</b>	10 (55%)	6 (67%)	4 (44%)
Soft mandates: <i>Fiscal Adjustment</i> (2003); <i>Capital Account Liberalization</i> (2005); <i>International Trade Policy</i> (2009)	3 (17%)	2 (22%)	1 (11%)
Activity management: <i>PRSPs and the PRGF</i> (2004); <i>Technical Assistance</i> (2005); <i>Jordan</i> (2005); <i>FSAP</i> (2006); <i>Multilateral Surveillance</i> (2006); <i>Interactions with Member Countries</i> (2009); <i>Research</i> (2011)	7 (39%)	4 (44%)	3 (33%)
<b>Total</b>	18 (100%)	9 (100%)	9 (100%)

As shown there, the largest sub-category is activity management studies, which have accounted for almost 40 percent of IEO evaluations.

### *Evaluation Investigations*

Evaluation investigations covered three areas—governance, financial crises, and operational policies. They generally started from the hypothesis that there was something amiss in Fund performance—with, for example, the *Exchange Rate Policy Advice* evaluation noting that it focused “deliberately on what [was] not working well”—and then set about assembling and analyzing evidence in order to accept/reject/refine that hypothesis, considering possible reasons for it, and identifying steps for correcting underlying problems.

Unique in its focus on “Management and above,” the *Governance* evaluation looked at the IMF’s accountability systems (or the lack thereof) for the Executive Board and Management. This evaluation was launched against the backdrop of widespread questioning of the IMF’s legitimacy and criticism, especially with respect to the small shares of emerging and other developing economies in Fund decision-making structures, relative to those of the advanced economies. A unique aspect of this evaluation was the fact that no Summing Up was issued following the Executive Board discussion; instead, the Managing Director and Board issued a joint statement embracing the report “as part of an ongoing process to strengthen the IMF’s governance framework” while also pointing to the complexity and interrelatedness of the various issues involved and the fact that addressing them would take time.<sup>5</sup> Three crisis evaluations (*Capital Account Crises*, *Argentina*, and *Financial and Economic Crisis*) investigated Fund

<sup>5</sup>See “Joint Statement by the Executive Board and the IMF Managing Director,” Press Release No. 08/121, May 27, 2008, available at [www.imf.org](http://www.imf.org).

performance in the run-up and/or response to major crises. These three evaluations spanned an eight-year period that saw the origins of major financial crises passing from emerging economies to advanced economies. The shift was reflected in the IEO's coverage—with 2003's *Capital Account Crises* evaluation, which focused on Brazil, Indonesia, and Korea, and 2004's *Argentina* evaluation, giving way to 2011's *Financial and Economic Crisis* evaluation, which focused on the United States and the United Kingdom. Broadly speaking, each of these evaluations analyzed (1) what happened before, during, and/or after the crisis; (2) what role the Fund played; (3) how the Fund might have performed better; and (4) what had prevented the Fund from performing better. This said, they differed in their coverage of these questions, with much greater attention given to the diagnostic question—what had prevented the Fund from performing better—in the 2011 crisis evaluation than in the 2003 and 2004 crisis evaluations.

Four policy-related evaluations (*Prolonged Use, Aid to Sub-Saharan Africa, Exchange Rate Policy Advice, and Structural Conditionality*) investigated the extent of and reasons for noncompliance with specific IMF operational policies. These evaluations spanned a six-year period, with *Prolonged Use* the IEO's first evaluation, issued in 2002, and the other three all issued in 2007. Each covered compliance with a specific IMF operational policy, though the roots of those policies varied—from an implied mandate in the case of *Prolonged Use* to explicitly Board-approved policies in the case of *Aid to Sub-Saharan Africa* and *Structural Conditionality* to the Fund's 1977 Surveillance Decision in the case of the *Exchange Rate Policy Advice* evaluation. Each of these evaluations explored the foundations of the governing policy, as a benchmark for assessing the degree of compliance deviations therefrom, and the reasons for any deviations.

### Evaluation Studies

IEO's 10 evaluation studies were typically undertaken with a more agnostic view about Fund performance than were the investigations, and fell into two broad sub-categories. One sub-category—containing *Fiscal Adjustment, Capital Account Liberalization, and International Trade Policy*—dealt with the Fund's practices and positions on “soft mandates”: substantive issues for which the institution did not have a clearly defined or Board-approved operational policy. The other sub-category—containing *PRSPs and the PRGF, Technical Assistance, Jordan, FSAP, Multilateral Surveillance, Interactions with Member Countries, and Research*—dealt with the execution and impact of Fund activities and how these might be improved.

The three “soft-mandate” evaluations took varying approaches, though documentation of IMF practice was central to all three. All were closer in spirit to the policy investigations than to the other evaluation studies, looking at what the institution did when it had no specific policy on the issue in question. *Capital Account Liberalization* and *International Trade Policy* gave explicit attention to the legal and operational policy foundations (or lack thereof) of Fund advice. *Capital Account Liberalization* took as a central premise the lack of an actual policy and/or Board decision; it looked at Fund advice on capital account liberalization during a period of shifting views that was characterized by IMF Management support for

an amendment to the Articles in the pre-East Asia crisis period, giving way to a more open position after the East Asia crisis erupted. *International Trade Policy* explicitly summarized the debate surrounding the legal and operational policy foundations of the Fund's work on trade, before setting that discussion aside and selecting as evaluation benchmarks whether the IMF's advice was well thought out, linked to macro policies, and evenhanded. *Fiscal Adjustment* took a similar approach in its selection of evaluation benchmarks: it judged programs' fiscal stance by their appropriateness relative to countries' underlying economic and financial situations. However, unlike *Capital Account Liberalization* or *International Trade Policy*, the *Fiscal Adjustment* evaluation made little mention of the governing IMF policy framework. Exceptionally, as discussed in the section "IEO's Findings and Recommendations" below, it analyzed poverty reduction and social protection issues arising in the context of Fund-supported programs, and recommended that the Fund delineate a clear operational framework for dealing with such issues.

The seven activity-management evaluations focused on how and how well the Fund carried out different activities. IEO selected these activities for study for a variety of reasons.

- Three of the activity management evaluations—*Jordan*, *Multilateral Surveillance*, and *Interactions with Member Countries*—focused on core IMF activities of multilateral surveillance and country relations. The latter two both covered bilateral surveillance, programs, and technical assistance but they had very different origins and scope. *Interactions with Member Countries* shared some of the motivational elements of an investigation, following up on concerns about relationship management that had surfaced in the *Exchange Rate Policy Advice* evaluation. By contrast, the *Jordan* evaluation had been launched as an IEO experiment (subsequently abandoned) in periodic examinations of more-or-less routine IMF work on country matters. The two evaluations also differed fundamentally in scope. In *Interactions with Member Countries*, the very broad all-member coverage provided a natural comparative framework for the evaluation, which focused on how IMF interactions varied across the different country groups, but that same breadth of coverage complicated the execution of the evaluation. The *Jordan* evaluation, in contrast, had the execution advantage of narrow focus, but as a single-country study it lacked built-in comparative benchmarks, forcing it to consider Fund performance against some absolute—albeit never explained—standard. In terms of scope, the *Multilateral Surveillance* evaluation had the best of both worlds: like *Jordan*, it was narrowly focused, easing its execution, but like *Interactions with Member Countries* it successfully established an internal comparative framework, looking at the Fund's *World Economic Outlook* on the one hand versus the *Global Financial Stability Report* on the other and also specifically comparing multilateral surveillance with aspects of bilateral surveillance.
- Two evaluations—*PRSPs and the PRGF* and *FSAP*—focused on IMF initiatives that were both relatively new at the time and that had been developed and executed in close partnership with the World Bank. Carried out in

parallel with evaluations by the World Bank's Operations Evaluation Department/Independent Evaluation Group, they assessed experience with the programs to date, focusing on the Fund's role and performance. They were very much akin to IMF staff progress reports on these initiatives, with their main value over such staff reviews being the independence of their perspective. They were launched in advance of Board reviews of the initiatives, and provided inputs into those reviews.

- The remaining two activity management evaluations, on *Technical Assistance* and *Research*, had limited operational policy content albeit significant operational relevance—*Technical Assistance* in direct and obvious ways and *Research* in that the evaluation's scope included the selected issues papers prepared for Article IV consultations. Like the *PRSPs and the PRGF* and *FSAP* evaluations, those of *Technical Assistance* and *Research* provided straightforward assessments of Fund performance, and identified ways of improving it. Both evaluations dealt with how the Fund decided on resource allocation and project selection within the particular activity (whether the overall technical assistance program or the overall research program). But neither addressed in any depth the issue of the appropriate amount of resources to be devoted to the activity in light of the IMF's comparative advantage relative to other providers of technical assistance and research an issue highlighted in the Lissakers Report, as discussed in the section "Evolution over Time" below.<sup>6</sup>

## Evaluation Evidence: Data and Methods

This section looks at the data and methods that IEO evaluations have used. It considers the approaches followed in the 18 evaluations, and briefly comments on the relative contributions of the various data sources and methods.

Qualitative sources and methods (interviews, IMF documents, and field visits) were used by all evaluations, while quantitative sources and methods (confidential surveys and regression or other empirical analysis) were used somewhat more selectively, in part because not all evaluations were amenable to such approaches. Where used, quantitative methods proved powerful in sharpening both the evaluation findings and the debate on them with IMF staff and Management. The handling of evaluation data and methods has been a frequent topic in IEO evaluation completion reports (ECRs), which are a self-assessment initiative launched in 2007.<sup>7</sup>

<sup>6</sup>The *Research* evaluation (para 2) did set out reasons why research is important to the Fund's credibility and contribution; however, like the *Technical Assistance* evaluation, it did not analyze the magnitude of the Fund's investment in research either relative to comparators or relative to possible alternative models involving more or less outsourcing.

<sup>7</sup>By end-2011, the IEO had produced evaluation completion reports (ECRs) for all evaluations since *Aid to Sub-Saharan Africa*. The main audience for ECRs is the IEO staff, but the reports are also available to the IEO's external evaluators. Most ECRs contain about 10–15 pages of text, discussing the strengths and weaknesses of the evaluation process and highlighting lessons learned by the team about evaluation execution and constraints the team faced. This is supplemented by extensive annex material detailing a number of issues, such as outreach, and notes on how particular issues were developed. ECRs contain a list of people interviewed for the evaluation and an inventory of where data and information is stored. These lists are confidential and to be shared only with the external evaluators of the IEO.

TABLE 8.4

Data and Methods in IEO Evaluations							
	Internal Documents	Interviews	Field Visits	Case Studies	Surveys	Regression Analysis	Expert Papers
<i>Governance</i>	•	•	•	•	•		•
<b>Crises</b>							
<i>Capital Account Crises</i>	•	•	•	•			•
<i>Argentina</i>	•	•	•	•			
<i>Financial and Economic Crisis</i>	•	•	•	•			•
<b>Policies</b>							
<i>Prolonged Use</i>	•	•	•	•	•	•	•
<i>Aid to Sub-Saharan Africa</i>	•	•	•	•	•	•	
<i>Exchange Rate Policy Advice</i>	•	•	•	•	•		
<i>Structural Conditionality</i>	•	•	•	•	•	•	
<b>Soft mandates</b>							
<i>Fiscal Adjustment</i>	•	•	•	•		•	
<i>Capital Account Liberalization</i>	•	•	•				
<i>International Trade Policy</i>	•	•	•	•	•	•	•
<b>Activities</b>							
<i>PRSPs and the PRGF</i>	•	•	•	•	•	•	
<i>Technical Assistance</i>	•	•	•	•		•	
<i>Jordan</i>	•	•	•	•			
<i>FSAP</i>	•	•	•	•	•		
<i>Multilateral Surveillance</i>	•	•	•		•		
<i>Interactions with Member Countries</i>	•	•	•	•	•		•
<i>Research</i>	•	•	•	•	•	•	•
<b>Total</b>	18	18	18	16	11	8	7

All 18 evaluations used internal IMF documents and interviews (of authorities, staff, and sometimes others), with field visits required in some instances to complete the interviews; other evaluation methods were used more selectively (Table 8.4 above). Sixteen evaluations used case studies. Eleven used confidential surveys of authorities and staff (and in some instances other stakeholders as well). Eight used regression analysis of IMF practice. Seven commissioned background papers by outside experts to provide perspectives on relevant external practice and/or to review and assess internal IMF work.

IMF documents were an important data source for IEO evaluations; publicly available documents were used by all evaluations and internal (not publicly available) documents were used by most.

- The most basic use of IMF documents common to almost all evaluations was for deep background, as IEO teams used them to familiarize themselves with how particular issues were viewed by the Executive Board and by IMF Management and staff.
- A second use of IMF documents was in comparing how the Fund had treated particular issues across large samples of countries, whether in the context of programs or surveillance. This approach was much less common but important in several evaluations, notably *Exchange Rate Policy Advice*, *Fiscal Adjustment*, and *Interactions with Member Countries*. Where successful, it enabled evaluations to, in effect, quantify underlying qualitative evidence about how particular issues were treated in Fund documents. But an important lesson learned from these evaluations is that the success of such approaches depends critically on the quality and consistency of the evaluation team's coding effort. This effort is greatly complicated when IEO coders have different backgrounds and degrees of familiarity with the issues under study. The use of Fund documents to compare experience across countries therefore carries major implementation risks, as highlighted in the Evaluation Completion Report for the *Exchange Rate Policy Advice* evaluation, which detailed the challenges of managing a team of coders tasked with making consistent judgments about the treatment of exchange rate issues in IMF documents.<sup>8</sup> Similar challenges surfaced in the evaluation of *Interactions with Member Countries*, which relied initially on a large team of coders of varying degrees of familiarity with IMF operations before shifting to a smaller and more cohesive team.
- A third use of IMF documents was in informing in-depth evaluation case studies, where internal documents—especially on the review process—provided a clear window onto the staff debate surrounding the Fund's approach to a particular issue. In *Interactions with Member Countries*, for example, the evaluation team's access to internal IMF staff memoranda enabled it to piece together what had happened behind the scenes many

<sup>8</sup>IEO Evaluation Completion Report—*IMF Exchange Rate Policy Advice, 1999–2005*,” September 2007.

years previously, despite contradictory interview reports about the associated events. And in some cases evaluation reports usefully cited specific internal memoranda, thereby providing Executive Directors and external audiences with highly relevant information about the internal staff debate that had surrounded the activities being evaluated. For example, in *Argentina*, such citations revealed the nature of the debate about whether briefing papers should include instructions for missions to discuss alternative policy frameworks, and thereby gave the reader a fuller picture of what had happened. But not all evaluation reports used this privileged source equally effectively. For example, the *Jordan* evaluation mentioned problems in Bank-Fund collaboration on public expenditure priorities and on conditionality regarding privatization, but went no further in explaining the nature of the problems, despite the team's access to the internal documents that told all.

All 18 IEO evaluations used interviews, and, as with IMF documents, in three different ways. But because the different ways were not always clearly spelled out in IEO evaluations, there is room for confusion about IEO protocols for recording the findings of exploratory interviews, and for reflecting them in evaluation reports.

- Exploratory interviews provided deep background and helped IEO staff to understand developments and formulate hypotheses both in preparing issues papers and throughout the evaluation process. In *Interactions with Member Countries*, for example, extensive interviews were carried out at the start of the evaluation, raising many questions that were then pursued via other evidentiary sources (such as the survey and field visits) for triangulation in the final report.
- Second, several evaluations aimed to quantify their interview results in terms of statements about “all,” “most,” or “many” interviewees’ views on particular issues. In these cases, the interview process was used as a variation on a survey process, with common questions asked across interviewees in structured or semi-structured interviews with pre-set questionnaires and careful recording of the results of the interviews as the embodiment of evaluation evidence. The *Governance* ECR highlights as a lesson learned the importance of taking a systematic approach to interviews, using structured questions that are sent to interviewees in advance, clear and timely minutes, and so on.<sup>9</sup> Clearly, such steps are important for evaluations that rely on the quantification of interview evidence.
- Third, in some other evaluations, extensive survey and/or empirical analysis lessened the reliance on interview evidence, except for purposes of interpreting or deepening the understanding of other evidence or for in-depth country case studies. In some cases interviews, just like documents, provided “smoking-gun” evidence—a feature especially relevant for investigation evaluations.

---

<sup>9</sup>“Evaluation Completion Report—*Governance of the IMF: An Evaluation*,” July 2009.

The 2002 evaluation report on *Prolonged Use*, for example, quoted from an interview with a senior Pakistani official to the effect that: “Most IMF-supported programs primarily supported political purposes. Thus it should come as no surprise that they did not achieve much in economic terms. . . .”

All 18 evaluations also used field visits, albeit with considerable variation in how these visits were pitched, conducted, and documented. Some field visits were part of in-depth evaluations’ country case studies (as for example in *Prolonged Use*, *Capital Account Crises*, *PRSPs and the PRGF*, *Argentina*, *Jordan*, *Aid to Sub-Saharan Africa*, and *Interactions with Member Countries*) and involved extended country stays and meetings with partners and stakeholders beyond the authorities. Country visits undertaken for some other evaluations (such as *FSAP* and *Exchange Rate Policy Advice*) were more narrowly focused, and mostly undertaken for the purpose of targeted meetings with key IMF interlocutors with whom IEO staff had not been able to connect during the Annual or Spring Meetings or by phone. Yet other country visits were in between. In terms of managing the interface with authorities and stakeholders in the context of field visits, IEO Management devoted considerable administrative effort to ensuring that the IEO appeared organized, relying on a centralized system for tracking staff interactions with country authorities and Executive Directors’ offices. Going forward, it would be useful for the IEO to develop a more decentralized and broadly accessible data base about past field visits and so on—mindful of confidentiality concerns where relevant—so that IEO staff planning future visits can do so more knowledgeably about prior IEO activities and, accordingly, more cost-effectively.

Country case studies were used in 15 evaluations, with issue-specific case studies used in a sixteenth evaluation. The depth and presentation of case studies varied widely, ranging from the *Argentina* and *Jordan* evaluations, which were themselves, in effect, case studies, to *Exchange Rate Policy Advice*, which identified 30 economies for “detailed analysis” of the dialogue on exchange rate issues; *Interactions with Member Countries*, whose case studies simply informed the treatment of the different country groupings in the main text; and *Research*, where the country (regional) case studies underpinned the analysis in the main report but are to be subsequently published. In between, IEO evaluations varied in the depth to which they developed their case studies and in the detail with which they presented their findings in reports. For example, *Prolonged Use* and *Capital Account Crises* both included lengthy case-study sections in the main reports; *PRSPs and the PRGF* summarized its case study results in a free-standing volume issued jointly by IEO and the World Bank’s evaluation group; *Aid to Sub-Saharan Africa* and *International Trade Policy* both included annexes briefly summarizing their case-study findings; and *Technical Assistance* included tables in its main text summarizing its assessment of the effectiveness of technical assistance in each case. On the whole, the retrospective found that country case studies can be an important qualitative tool, especially for supporting and complementing quantitative methods such as surveys and regression analysis. Used in this way they can deepen understanding of empirical results—both central values and outliers—as for example in the *Aid to Sub-Saharan Africa* and *Technical Assistance* evaluations, as discussed below.

Surveys were used in 11 evaluations, with their use increasing over time.<sup>10</sup> A frequent topic in ECRs, surveys are the data source that has most benefited from learning over time. IEO's more recent survey use has built on the survey experience gained by earlier IEO teams with respect to the management of survey design, responder interface, and the selection of contractors for survey execution.<sup>11</sup> Importantly, the growing popularity of surveys has reflected their ability to put numbers on qualitative issues—a very important feature in concretizing debates about performance, especially in a numerate staff culture like the IMF's. Evaluation surveys have enabled fruitful sets of comparisons, for example across country types, or between staff, authorities, and partners, that in turn have allowed the data to stand on their own, without requiring an absolute benchmark for judging whether the favorable (or unfavorable) responses were “high” or “low.”

IEO's survey evidence has proved a powerful tool for the evaluations that have used it—especially to *Aid to Sub-Saharan Africa*, *Exchange Rate Policy Advice*, *Interactions with Member Countries*, *Governance*, and *Research*—enabling comparative statements that advanced understanding of the issues under study in important ways. In *Aid to Sub-Saharan Africa*, for example, survey evidence was critical in establishing the relative harmony between IMF staff and authorities' views and the relative disharmony between IMF staff views and those of partners, such as World Bank staff, donors, and civil society. It also highlighted how staff views on the IMF's treatment of the Millennium Development Goals and related PRSP issues differed from the spin that the Fund's External Relations Department was putting on the same issues. In *Exchange Rate Policy Advice*, survey evidence showed how the large emerging economies held the Fund's exchange rate analysis and advice in much lower regard than did other country groups, especially the low-income countries. This finding found reflection across a number of issues in the survey work done for *Interactions with Member Countries*, which identified major disconnects between the views on Fund performance held by the authorities of advanced and emerging economies and by the IMF staff working on these countries. In *Governance*, the survey evidence revealed Executive Directors' views about the Board's limited expertise on financial management and other issues that they saw as important, and the even lower regard that the surveyed senior staff had of Board members' competence in this and other areas. In *Research*, the survey evidence showed that many authorities felt that IMF research was message-driven and that a majority of staff felt that their own research and its conclusions needed to be aligned with Fund views.

Eight evaluations used regression analysis, four of them in the first part of the retrospective period and four in the second.<sup>12</sup> Among the eight, the analysis was decisive in producing key findings in three: *Fiscal Adjustment*, *Technical Assistance*,

---

<sup>10</sup>Among the 11, *Prolonged Use* used a written questionnaire managed by the IEO.

<sup>11</sup>For example, “Evaluation Completion Report—An IEO Evaluation of IMF Involvement in International Trade Policy Issues” 2009.

<sup>12</sup>Empirical analysis was also developed for *Multilateral Surveillance*; however, the evaluation report did not include it.

and *Aid to Sub-Saharan Africa*. In those three, IMF operational data were used to analyze Fund practice and the empirical results contributed to the evaluations' core findings. In *Fiscal Adjustment*, the regressions established the facts about trends in program-supported fiscal corrections. In *Technical Assistance*, they showed the disconnect between IMF technical assistance programs and country priorities as proxied by country poverty reduction strategies.<sup>13</sup> In *Aid to Sub-Saharan Africa*, they related countries' starting positions on inflation and external reserves to PRGFs' programmed spending and absorption of aid, and in so doing estimated the magnitudes of key parameters of Fund practice. In the other evaluations in which regression analysis was used, it also made a contribution even if it was not the decisive evidentiary plank: *Prolonged Use* analysis identified the key characteristics of prolonged users compared to intermittent users; *PRSPs and the PRGF's* cross-country analysis showed how PRGFs generally targeted smaller fiscal adjustment than earlier programs; *Structural Conditionality* analysis showed that compliance with structural conditionality had little impact on sectoral outcomes; *International Trade Policy* analysis found only weak evidence of a favorable effect of trade conditionality on actual trade flows; and *Research* found broadly similar results for the citation and publication of IMF research compared with those of research by other international institutions and by central banks.

Externally authored background papers were used in seven evaluations, mostly in the second half of the retrospective period.<sup>14</sup> These papers took two forms.

- The first were expert papers that the IEO commissioned to bring in relevant external perspectives and/or credibility. Such papers made a major contribution to the *Governance* evaluation, for which qualitative and comparative analysis loomed large in the evidentiary base. For *International Trade Policy*, the IEO commissioned expert papers to assess IMF performance on specific issues, such as trade in financial services and preferential trade arrangements. *Interactions with Member Countries* also drew on external expertise to bring in fresh perspectives, for example on Fund interactions with emerging economies, civil society organizations, and parliamentarians.<sup>15</sup>
- The second type of externally authored papers were consultant reviews of evaluation evidence. For the *Financial and Economic Crisis* evaluation, the IEO commissioned external reviews of the Fund's pre-crisis publications to

<sup>13</sup>The evaluation, however, did note the finding of correlation between IMF technical assistance support and programs supported by the Poverty Reduction Growth Facility and the Extended Fund Facility.

<sup>14</sup>Additional evaluations, including *Fiscal Adjustment*, *FSAP*, *Multilateral Surveillance*, and *Aid to Sub-Saharan Africa*, have utilized team-prepared background papers, some of which have been issued by the IEO and/or other IMF units.

<sup>15</sup>Of course, not all the work commissioned by the IEO ended up in authored background papers. For example, in *Aid to Sub-Saharan Africa*, an outside expert's contribution ended up simply informing relevant parts of the evaluation and its revisions rather than being issued as a freestanding paper, and in *Interactions with Member Countries* at least one commissioned paper was put aside as, ultimately, its topic was not covered in the evaluation report and there were concerns that the paper's conclusions might be construed as endorsed by the IEO.

identify banner messages inter alia, and for the *Research* evaluation, external consultants assessed the quality of IMF selected issues papers and working papers. In some cases, such reviews have not differed materially from team-prepared reviews, but being explicitly labeled as externally authored has afforded them somewhat greater editorial distance from the IEO and hence greater freedom to express their authors' opinions. Such background papers do not enjoy absolute freedom, however, because the IEO's ownership of these papers (and its associated potential reputational risk) is not zero.

## IEO's Findings and Recommendations

Findings and recommendations are the twin pillars of all evaluations, and the 18 IEO evaluations considered in this paper are no exception. This section considers their presentational aspects and content, and then the connections between them across individual evaluations.

### Findings in IEO Evaluations

IEO's presentation of findings generally took the form of a narrative, rather than the enumerated lists that were more common in its presentation of recommendations.<sup>16</sup> Rough counts are possible to convey orders of magnitude; they tend to number between 5 and 10 findings per evaluation, depending on how finely or roughly the findings were packaged.

Several recurring themes permeated IEO's evaluation findings—in some cases reflecting follow-up on previous evaluation work. The three most frequent findings concerned:

- ambiguity and confusion (among IMF stakeholders and staff) about the IMF's governing policies or mandates—as observed in the policy and soft-mandate evaluations;
- lack of candor in IMF staff reports—as observed in the crisis and country-based evaluations; and
- limited coordination between the Fund's macroeconomic and financial sector analysis—as observed in the evaluations of crises, *Exchange Rate Policy Advice*, *Multilateral Surveillance*, and *FSAP*—and, relatedly, limited IMF coverage of macro-financial sector linkages, as observed in the *Research* evaluation.

These recurring findings also found reflection in recurring recommendations, as discussed later in this chapter.

---

<sup>16</sup>Most of the IEO's 18 evaluations contain a prominent section labeled "Findings." The others also contain findings but headline them differently. For example, *Prolonged Use* and *Interactions with Member Countries* label their findings "Conclusions;" *Structural Conditionality* and *Research* label theirs "Findings and Conclusions" or "Conclusions and Findings," *Capital Account Crises* labels its findings "Assessment," and *FSAP* labels its findings "Lessons." The point here is simply that in all 6 of these cases, the existing content would be as suited to a label of "findings" as that in the 12 evaluations that use this term. This section therefore treats them all as findings.

*Governance* was the only evaluation of its kind but one of its key findings—about the Board’s focus on executive rather than supervisory functions—provides a powerful lens for viewing the findings of other IEO evaluations. This is especially the case for the policy and soft-mandate evaluations, which found that member countries received variable treatment in their interactions with the Fund, including in the implementation of operational policies, all with Board concurrence. The key point here is that with a highly involved Executive Board, which approves both policies and decisions about the implementation of those policies in individual country cases, the scope for case-by-case approaches is enhanced, bringing to IMF interactions with countries the benefits of a tailored approach but also the risks of uneven treatment. As reflected in IEO’s policy and other evaluations, the use of case-by-case approaches has sometimes created or exacerbated confusion among IMF stakeholders and staff about what Fund policy actually was on particular issues.

The three crisis evaluations found that over-optimism and lack of candor—born of what was seen with hindsight to have been undue concern to maintain good country relations—had contributed to the Fund’s poor performance in the run-up and/or response to crises. The facts underlying the crises differed, ranging from the public or private sector origins of the crises in Korea, Indonesia, and Brazil to the questions about Argentina’s exchange rate policy and exit strategy to the particulars of the post-Lehman-collapse global financial crisis. But the three evaluations share common ground on the reasons for the Fund’s failure to better anticipate each crisis and/or to deal with it once it struck. Each evaluation highlighted analytical weaknesses, organizational impediments, internal governance problems, and political constraints related to concerns about country relations.

In each of the four evaluations of operational policy, IMF policy ambiguity and uneven policy implementation across countries, reinforced by mixed signals from the Board, were central diagnostic findings. Like the crisis evaluations, each policy evaluation offered factual findings on the particulars under review. Thus *Prolonged Use* found that the use of Fund resources had increased in line with various “demand-side” factors that were reinforced by internal Fund cultural conditions related to over-optimistic forecasts, lack of candor, and political constraints. *The IMF and Aid to Sub-Saharan Africa* found mixed implementation of relevant policies, with more conformity with some aspects of Fund policy and less with others. In both cases, this evaluation found messages being communicated by the Fund externally to be at variance with the reality on the ground—aggravating an already confused situation about what the Fund stood for in Africa. *Exchange Rate Policy Advice* found unclear rules of the game, uneven focus on factors driving exchange rate developments, and a lack of operational guidance on key issues such as exchange market intervention. *Structural Conditionality* found that no changes had taken place in the volume of structural conditionality, notwithstanding the Fund’s attempt to limit such conditionality through the streamlining initiative introduced several years earlier; it did, however, find a shift in the composition of such conditionality in the direction of the Fund’s core competences. These four evaluations also shared several diagnostic findings about the reasons for the Fund’s departures from full compliance. All of these findings implied a profound lack of

clarity on the underlying policy and its operational implications, born of the apparent differences of opinion on key aspects among Executive Directors.

The three soft-mandate evaluations also found uneven implementation of IMF policies across countries. One of them—*Fiscal Adjustment*—highlighted this as a positive finding, refuting Fund critics’ characterization of the institution’s approach as “one-size-fits-all.”<sup>17</sup> But a common finding of the three evaluations was that the cross-country differences did not reflect the systematic and consistent tailoring to country conditions of a clearly articulated approach; rather they simply reflected differences that were typically unexplained in Fund documents, because there was no policy benchmark against which to gauge the degree of implementation. The three evaluations did not fully explore these cross-country differences in treatment. Instead they focused more on the underlying substantive issues, although the *Capital Account Liberalization* evaluation was able to associate country differences in IMF treatment with differences in country views and preferences—and in IMF staff views. Going forward, these features suggest that IEO evaluation studies of soft mandates could usefully focus much more on how, in the absence of Board-approved guidance, the Fund formulates its positions on particular issues and the extent to which countries receive evenhanded treatment with regard to these issues.

In line with their orientation as studies, the findings of the activity evaluations all tended to focus on the quality, relevance, and effectiveness of their highlighted activities. All found evidence of supply-driven “silo” approaches—though not all of them used that term—with the specialist groups that championed the various programs much more behind the programs than were other Fund staff. Of course, some silo-ism is unavoidable in a complex institution, and can be very efficient. So a key question is really whether the IMF had too much or too little of it. This question has not been explored in IEO evaluations to date but should be in future, especially in activity evaluations. Also noteworthy are the recurring findings of: (1) weak traction with member countries, in the evaluations of *Multilateral Surveillance* and *Interactions with Member Countries*; (2) weak influence of the PRSP process and the FSAP initiative on the operational work of area department staff; and (3) limited attention to inputs from member countries in the design of country programs, in the *Technical Assistance* evaluation, and of research projects, in the *Research* evaluation.

## Recommendations in IEO Evaluations

This section briefly considers the number and content of IEO recommendations.<sup>18</sup>

<sup>17</sup>To some extent, this finding illustrates how these evaluations were the mirror image of policy evaluations; rather than looking at whether and how IMF policies were being implemented (or not), they examined whether and how “non-policies” were being implemented.

<sup>18</sup>All but one of the 18 evaluations contained a prominent section labeled “Recommendations.” The exception was the *Jordan* evaluation, which contained a section labeled “Lessons Learned” (see *IMF Support to Jordan: 1989–2004*, 2005; p. 3). As compared with the other 17 evaluations’ recommendations, these lessons were broadly similar in nature, though they were worded a little differently. In presenting the lessons, the evaluation report noted that they were not “couched as recommendations” as they had more general applicability beyond the Jordan program. For ease of presentation, this paper treats the *Jordan* evaluation lessons as recommendations.

TABLE 8.5

## Numbers of Recommendations in IEO Evaluations

	Number of Headline Recommendations	Number of Sub- Recommendations		Total: Headline and Sub- Recommendations ("Should's" Not "Could's")	Pages Covering Recommendations
		"Should"	"Could"		
<i>Governance</i>	4	25	0	29	2½
<b>Crises</b>					
<i>Capital Account Crises</i>	6	23	0	29	4
<i>Argentina</i>	6	10	1	16	3
<i>Financial and Economic Crisis</i>	5	0	19	5	2½
<b>Policies</b>					
<i>Prolonged Use</i>	14	13	0	27	8
<i>Aid to Sub-Saharan Africa</i>	3	5	0	8	½
<i>Exchange Rate Policy Advice</i>	11	6	3	17	4
<i>Structural Conditionality</i>	6	8	1	14	1
<b>Soft mandates</b>					
<i>Fiscal</i>	5	6	4	11	2¼
<i>Capital Account Liberalization</i>	2	0	3	2	1½
<i>International Trade Policy</i>	6	9	0	15	1¼
<b>Activities</b>					
<i>PRSPs and the PRGF</i>	6	23	0	29	4½
<i>Technical Assistance</i>	6	10	10	16	2½
<i>Jordan</i>	9	0	0	9	1
<i>FSAP</i>	7	12	12	19	6
<i>Multilateral Surveillance</i>	4	8	18	12	3
<i>Interactions with Member Countries</i>	9	0	0	9	3
<i>Research</i>	8	0	0	8	½
<b>Total</b>	117	158	71	275	48½
<b>Average</b>	6½	9	4	15	2½

*Numbers of Recommendations*

IEO's 18 evaluations contained 117 headline recommendations and 158 sub-recommendations, for a combined total of 275 (Table 8.5). These numbers are, however, a small exaggeration, because some recommendations were made in more than one evaluation and are thus double counted in the simple tally (Box 8.1). When the recurring recommendations are counted only once, the headline total falls to 104 and the overall total (headline recommendations plus sub-recommen-

## BOX 8.1

### Recurring Headline Recommendations

Six headline recommendations recurred in two or more evaluations.<sup>1</sup>

- The most popular recommendation, recurring in five evaluations (*Prolonged Use, Aid to Sub-Saharan Africa, Exchange Rate Policy Advice, Structural Conditionality, Capital Account Liberalization*) was for clarification of the respective IMF policies and/or positions. Each of the five evaluations called for clarification of the policies in its own area of focus.
- Two recommendations recurred in four evaluations. These concerned the need for greater:
  - Candor about downside risks (*Capital Account Crises, Financial and Economic Crisis, Prolonged Use, Jordan*); and
  - Integration of macroeconomic and financial sector analysis (*Financial and Economic Crisis, Exchange Rate Policy Advice, FSAP, Multilateral Surveillance*).
- Three recommendations recurred in two evaluations. These concerned the need for greater attention to:
  - Country political economy underpinnings (*Prolonged Use and Jordan*);
  - Country dialogue (*Exchange Rate Policy Advice and Interactions with Member Countries*); and
  - Monitoring and evaluation (*Aid to Sub-Saharan Africa and Structural Conditionality*).

<sup>1</sup>One sub-recommendation was made in two evaluations. It was on Board Summings Up, and appeared in the *Governance* and *Financial and Economic Crisis* evaluations.

dations) falls to 261—giving an average of 14½ headline and sub-recommendations per evaluation. In addition, there is not total clarity in all cases about what the IEO intended as headline and sub-recommendations, as different evaluations presented their recommendations differently.<sup>19</sup> Nevertheless, there are three key points here. First, these simple counts—whether corrected for double counting and ambiguities or not—provide a starting point for a discussion, in particular by providing a rough metric for comparing the evaluations under study. Second, in so doing, they also provide an order-of-magnitude indicator of the amount of work and resources involved for Fund staff in following up on IEO recommendations. This leads to the final point: whether the total number is 261 rather than 275—or even 200—these are big numbers if they are meant to entail serious follow-up.

The numbers of recommendations contained in IEO evaluations varied widely (Table 8.5). Eight evaluations made relatively few headline recommendations (2–5 each) and two made relatively many (11–14), with eight evaluations falling in

<sup>19</sup>These numbers cited in the text and set out in Table 8.5 reflect earlier attempts to reconcile the various counts for recommendations with counts in the parallel IEO exercise on recommendations. See Louellen Stedman, “IEO Recommendations: A Review of Implementation,” Chapter 9 in this volume.

between (6–10). Taking headline and sub-recommendations together, five evaluations made 1–10; nine evaluations made 11–20; and four evaluations made 21–30.

How should these variations in number be understood? Some insight can be gained by looking at those evaluations that provided the largest numbers of headline recommendations: *Prolonged Use* and *Exchange Rate Policy Advice*, with 14 and 11, respectively. Comparison of their recommendations with those of the other two policy evaluations, *Aid to Sub-Saharan Africa* and *Structural Conditionality*, which provided, respectively 3 and 6, suggests major differences in the degree of detail. Though the differences partly reflected differences in style among the lead authors of the four reports, there are also major differences between *Prolonged Use* and *Exchange Rate Policy Advice*. *Prolonged Use* contained a number of new and concrete proposals on a *broad* set of themes, such as on signaling; the need for treating prolonged use by aid-using low-income countries differently from that by other countries; the concept of selectivity; the desirability of ex post assessments; and so on—all of which emerged from the evaluation’s diagnosis of why prolonged use had proved so enduring. In contrast, the recommendations of *Exchange Rate Policy Advice* mostly involved variations on a relatively *narrow* theme—related to the need for improved exchange rate analysis and greater assurances of confidentiality as elements in improving the effectiveness of the Fund’s exchange rate policy advice.

Even wider variation is seen in the numbers of sub-recommendations, which ranged from 0–5 in six evaluations to 23–25 in three others: *Governance*, *Capital Account Crises*, and *PRSPs and the PRGF*. A brief examination of these three evaluations suggests that they simply contained a lot of detail, which then raises questions for follow-up to IEO recommendations: Does the IEO intend that each and every one of its sub-recommendations has the same standing as its headline recommendations?

- The *Governance* evaluation’s 25 sub-recommendations provided specificity for each of its headline recommendations. In positioning the sub-recommendations, the evaluation indicated that it “propose[d] detailed measures specific to the IMFC, the Board, and Management.” And while the text occasionally used the term “could” in describing how the proposals might work, they appear to have been intended as sub-recommendations rather than merely ideas for consideration.
- *Capital Account Crises* provided 23 sub-recommendations under its 6 core recommendations. Each was quite specific, and the language conveys that they were bona fide sub-recommendations, rather than ideas or suggestions.
- The *PRSPs and the PRGF* evaluation also provided 23 sub-recommendations. In introducing them, the evaluation stated that it made 6 broad recommendations, setting out directions for change and some ideas rather than a blueprint. However, the 23 sub-recommendations were fairly specific and prescriptive and in some cases they were complemented by additional ideas on how they might be implemented.

Half of the evaluations also explicitly included suggestions for consideration. These suggestions are not counted in the above totals, and they are listed in

Table 8.5 in the column labeled “could.” Four evaluations—*Technical Assistance*, *FSAP*, *Financial and Economic Crisis*, and *Multilateral Surveillance*—account for 70 percent of these suggestions, as discussed below.

- The *Technical Assistance* evaluation provided a total of 20 items that might be considered sub-recommendations. The first 6 were clearly introduced as “suggestions” that might be considered in implementing the evaluation’s recommendation to establish a medium-term country policy framework for technical assistance. Of the remaining 14, 4 were worded tentatively, using terms such as “could” rather than “would;” whereas the other 10 were clearly worded as recommendations rather than ideas or suggestions.
- *FSAP* contained 24 bulleted items that might be considered sub-recommendations. But of these, only 12 called for the IMF to take certain actions.<sup>20</sup> The other half were labeled steps that “could” be considered, rather than actions that “should” be implemented.
- *Financial and Economic Crisis* contained 19 subsidiary recommendations that it put forward as “. . . more specific suggestions on how [the five general recommendations] could be implemented. These specific suggestions should be seen as a starting point for further reflection; they are not necessarily the only way to follow through, and alternative approaches could have significantly different resource implications. . . .”
- *Multilateral Surveillance* contained eight bulleted subsidiary recommendations, with 18 additional suggestions that provided possible steps or options for implementing these subsidiary recommendations.

### Content of Recommendations

Though most IEO recommendations were process-related, calling for changes in the way things were done within the Fund, there were exceptions. Notable here are the recommendations on substantive content that were provided by the three soft-mandate evaluations. *Fiscal Adjustment* recommended the Fund to delineate an operational framework for addressing social issues, following the evaluation’s analysis of social spending in programs, though it did not consider the policy consistency of such a recommendation. *Capital Account Liberalization* recommended the Fund to pay greater attention to the supply side of capital movements in its surveillance activities. And *International Trade Policy* called for greater attention to preferential trade agreements and trade in financial services as part of surveillance and program activities. This pattern of calling for substantive changes contrasts with the approach that was taken in most of the policy evaluations, which typically sought greater clarity of the governing policy—so as to facilitate evenhanded implementation—but without taking a position on what that policy should be.

<sup>20</sup>Two of these sub-recommendations are amplifications of other sub-recommendations, describing how their implementation might be tailored to particular circumstances.

Broadly speaking, most recommendations flowed from the evaluations' findings and in turn from the evaluations' questions and evidence. Visual inspection of the findings and recommendations shows their alignment, as for example in the three crisis evaluations where the links between the findings and recommendations were clearly drawn. But it also reveals some looser connections. *Multilateral Surveillance*, for example, based its findings mainly on the evidence it presented about *technical* quality and *internal* coordination and related production challenges. But it focused its recommendations on the *strategic* uses of multilateral surveillance outputs, especially with respect to possible engagement with high-level *external* players. And in *Structural Conditionality*, most of whose recommendations drew on the technical analysis underpinning the findings, one was not linked to the evaluation findings or evidence. Beyond these outliers there is the issue related to the sub-recommendations: many of these appear to have been just the ideas of the evaluators no matter how they were labeled, with the particulars reflecting as much the creativity and experience of the evaluation team as the objective issues of the evaluation.

Finally, not all the evaluations used a logical framework to explain how the IEO moved from evaluation findings to evaluation recommendations.

- Five evaluations—*Argentina*, *Capital Account Liberalization*, *PRSPs and the PRGF*, *Aid to Sub-Saharan Africa*, and *Financial and Economic Crisis*—had useful sections that explicitly bridged from their findings to their recommendations. In two cases—*Argentina* and *PRSPs and the PRGF*—these sections were labeled “lessons learned,” and in the other three cases (*Capital Account Liberalization*, *Aid to Sub-Saharan Africa*, and *Financial and Economic Crisis*) they were labeled differently.
  - In *Argentina* and *Capital Account Liberalization*, the bridge sections covered the implications of the more factual findings, thereby setting the scene for the recommendations to come.
  - The bridge section of *PRSPs and the PRGF* set out the implications of the findings as well as diagnosing some of the reasons for them.
  - The bridge sections of *Aid to Sub-Saharan Africa* and *Financial and Economic Crisis* focused on the causes of the problems identified in the findings, with diagnostic chapters respectively entitled “Institutional Drivers of IMF Behavior” and “Why Did the IMF Fail to Give Clear Warning?” These sections segued directly into the recommendations.
- In the other 13 evaluations, the intermediate steps were either less explicit or less explicitly distinguished from the findings themselves.
  - In 8 of the 13 reports—*Prolonged Use*, *Capital Account Crises*, *Jordan*, *Fiscal*, *FSAP*, *Technical Assistance*, *International Trade Policy*, and *Exchange Rate Policy Advice*—the evaluations' assessments or lessons were interwoven with the findings.
  - In the other 5 reports—*Multilateral Surveillance*, *Structural Conditionality*, *Research*, *Governance*, and *Interactions with Member Countries*—the findings are presented quite nakedly, with the text proceeding immediately from them to recommendations.

## Evolution over Time

The IEO has evolved over time, shaped by a number of factors, including importantly the changing external environment in which the IMF and the IEO operate, feedback from the Executive Board's Evaluation Committee and other stakeholders, and changing IEO Management, staff members, and lead evaluation authors. These factors have driven the changes highlighted earlier between the IEO's first five years and its second five years, including (1) the shift in country coverage to more all-member evaluations; (2) the shift to investigations—especially on policies—away from studies; (3) the increased use of surveys and commissioned papers as evidentiary sources; and (4) the reduction in the number of main recommendations.

An additional influence on the IEO's evolution was the external evaluation of the IEO at its five-year anniversary—the so-called Lissakers Report, which assessed the IEO's performance in the first five years and provided a number of recommendations for change. Several of these recommendations pertained to staffing, external dissemination, and the cultivation of external constituencies for the work of IEO—topics that are not discussed here, as they go beyond the scope of this chapter. But three others are relevant here, and for ease of reference are reproduced in Box 8.2. In brief they are:

- Have more “bite”: don't neglect country cases and other sensitive topics.
- Focus on “why” questions when something goes wrong or to explain IMF involvement, and not just on “what” questions about IMF processes and procedures.
- Shorten and simplify: target IEO reports on more senior and broader audiences.

How well has the IEO been meeting these recommendations since the Lissakers Report was issued in 2006?

Has the IEO sharpened its bite? To address this question, the author compared pre- and post-Lissakers evaluations for (1) the depth of important adverse findings; (2) the unbundling of responsibility for any such adverse findings among the Board, Management, and/or staff; (3) the tonality with which any adverse findings were presented; and (4) the degree of positive statements about Fund performance. The comparison suggests that on average there was a modest increase in “bite” between the pre- and post-Lissakers evaluations.<sup>21</sup> It also suggests that the increase reflected three factors, each of which is consistent with the Lissakers report's recommendations.

- First, as noted earlier, the balance of evaluations shifted, towards relatively more investigations and relatively fewer studies in the post-Lissakers period. Since investigations generally have more bite than studies, it is not surprising that bite increased on average.

<sup>21</sup>The bite ratings are based on comparative readings of the evaluation reports, staff comments, and Summings Up, with the reports judged on their degree of criticism (explicit or implicit) of the institution's professional competence, independence, and/or evenhandedness. For the most part this refers to staff, the exception being with respect to the Board on governance.

## BOX 8.2

### Lissakers Report on IEO Evaluations

The Lissakers Report made several observations on IEO evaluations. The three paragraphs from its executive summary most relevant to this paper are reproduced below.

“Careful topic selection is vital, given the IEO’s limited resources. There are strong pressures pushing the IEO in the direction of evaluating broad subjects and staying away from areas, especially individual country cases, deemed sensitive by IMF management or member governments. The IEO should resist these pressures. Country programs are where IMF policies hit the ground and are tested and where the stakes are highest. Heightened sensitivity reflects their importance. Close examination of country cases can shed light on broader systemic issues and the IEO should not shy away, even where programs are on-going. To be effective, a watchdog must have a bite.”

“IEO evaluations to date are generally considered of high quality, but several criticisms were repeatedly made to the panel: they do not isolate and analyze in depth the most important questions such as why the IMF misdiagnoses exchange rate trajectories and over-estimates growth, nor do they tackle strategic institutional questions such as the IMF’s role in low income countries or why should the IMF (as opposed to other agencies) be engaged in technical assistance. The analyses instead focus heavily on IMF processes and procedures. The panel recommends a different mix of evaluators, greater use of peer review, and sharpening the IEO’s Terms of Reference to make clear its systemic role.”

“The panel agrees with the many who complain that IEO reports are too long and are becoming indistinguishable from other IMF documents, using the same terminology and the same frame of reference. IEO recommendations suffer the same weakness. This is not just a matter of readability. Making reports shorter and punchier is a way of forcing evaluators to be selective rather than comprehensive, to focus on the most important issues and to offer an analysis that will provoke thought well beyond the IMF staff and management. More disciplined reports will lead to more pointed recommendations.”

Source: “Report of the External Evaluation of the Independent Evaluation Office,” March 29, 2006, available at [www.imo-imf.org/ieo/files/evaluationofieo/032906.pdf](http://www.imo-imf.org/ieo/files/evaluationofieo/032906.pdf).

- Second, in identifying and explaining performance issues, there was a pronounced trend towards unbundling in the second period, as the IEO moved from accountability statements about “the IMF” in general to more pointed statements about the Board, Management, and/or staff.
- Third, pair-wise comparisons of individual pre- and post-Lissakers evaluations suggest that a more critical tone was taken in the later period by some—although not all—individual evaluations.
  - Consider the comparison of *Prolonged Use* (pre-Lissakers) with *Exchange Rate Policy Advice* (post-Lissakers). Like all the policy evaluations, both found substantial shortfalls in implementation. However, the tone of *Prolonged Use* was more positive and less normative than the tone of *Exchange Rate Policy Advice*. *Prolonged Use* focused on why the governing policy with respect to the Fund’s provision of temporary support had not

been implemented and how that lapse might be remedied through new approaches that dealt with the authorities', creditors', and donors' "demand" for prolonged use. *Exchange Rate Policy Advice* focused on the finding that the Fund's policy advice was not "as effective as it needed to be" and how that finding reflected Management and senior staff failures to ensure the appropriate "supply" of exchange rate analysis.

- However, not all post-Lissakers evaluations were more negative than their pair-wise comparators. For example, the post-Lissakers *International Trade Policy* evaluation praised the staff papers that had been recently issued on trade, just as its pre-Lissakers comparator *FSAP* praised the recent staff implementation of the FSAP initiative; and in both cases IMF staff responded positively to the evaluation. Similarly, both the pre-Lissakers evaluation of *Technical Assistance* and the post-Lissakers evaluation of *Research* took constructive approaches to setting out their findings and recommendations for change, and both were well received by staff.

Has the IEO focused more on "why" questions since the Lissakers Report? Based on the author's ratings of pre- and post-Lissakers evaluations for their coverage of internal governance, culture, and incentive issues, there was indeed an overall increase in attention to "why" and other diagnostic questions in the post-Lissakers cohort. To a large extent, this finding simply reflects the change in composition of IEO evaluations discussed above; policy investigations were more numerous in the post-Lissakers period and they naturally involved questions about why policies were not being implemented. Indeed, *Prolonged Use* merits a higher "why" rating than its post-Lissakers pair-wise comparator *Exchange Rate Policy Advice*, reflecting *Prolonged Use's* focus on the "demand-side" factors driving the variance between policy and practice, compared with *Exchange Rate Policy Advice's* focus on "supply-side" factors. Similarly, *Research* merits the same "why" rating as its comparator, *Technical Assistance*—whose approach Lissakers had faulted for not considering why or how much the IMF "as opposed to other agencies" should be involved in service delivery—because it took as given the amount of research the Fund carried out.<sup>22</sup>

Have the IEO's reports become shorter, more disciplined, and selective? Page lengths have shrunk dramatically: the average number of pages of IEO main reports (excluding annexes, references, executive summaries, and so on) fell by more than half, from 58 pre-Lissakers to 27 post-Lissakers. In accommodating the shortening, the IEO has experimented with approaches such as: (1) simply dropping some material, as in *Multilateral Surveillance*; (2) including large annexes in the main volume, as in *Aid to Sub-Saharan Africa* and *International Trade*; (3) including diskettes with additional material in the printed volumes, as in *Structural Conditionality, Governance, and Financial and Economic Crisis*; (4) posting background papers on the IEO website, as in *Interactions with Member Countries* and a number of other evaluations; and (5) publishing

<sup>22</sup>The *Research* evaluation report did express the view that the IMF should undertake at least some research in-house.

companion volumes with supplementary material, as in *Governance*. No doubt different readers will differ on their preferred packaging, but clearly it is time for the IEO to decide on what will be its signature approach. In all cases, a strong but brief executive summary should clearly set out the main findings and recommendations, something that was missing from one post-Lissakers report.

## Retrospective Conclusions

Two kinds of conclusions emerge from the above. First are those for the IMF and second are those for the IEO. They are addressed in turn below.

### For the IMF

For the IMF, the 18 evaluations taken together suggest three major conclusions. These reflect ongoing challenges within the IMF with respect to: (1) how it carries out its core mandates on international financial stability and surveillance; (2) how it interfaces with members; and (3) how its staff work, both among themselves and with others.

- On *core mandates*, IEO evaluations have repeatedly emphasized the need for greater IMF candor; better down-side risk analysis; and closer links between the Fund's macroeconomic and financial sector work.
- On *member interface*, successive IEO evaluations have identified departures from evenhandedness and the need for (1) greater transparency about cross-country differences in treatment; (2) more rules-driven approaches that are less political and not overly responsive to country relations concerns; and (3) greater clarity on approaches, policies, and follow up.
- On *modalities for IMF staff work*, IEO evaluations have consistently highlighted the need for more outward focus on members and less inward focus on staff; greater analytic independence and professionalism; and more cooperation and less silo-ing across units.

### For the IEO

For the IEO, the 18 evaluations taken together suggest five main conclusions. These pertain to work program design and work program execution, as set out below.

#### *IEO Work Program Design*

- *IEO's country coverage*. The IEO achieved greater balance across country groupings in its second five years, especially in its evaluations of IMF work with advanced economies. But with the exception of the important *Financial and Economic Crisis* evaluation, it did so largely indirectly, that is through the use of all-member evaluations, rather than as the result of an explicit risk-weighted approach. Going forward, it will be important for the IEO to pay relatively more attention to IMF work with the advanced

economies—beyond all-member reports—given the demonstrated high global costs of surveillance failures in advanced economies.

- *IEO's evaluation orientation.* As noted above, there was a rebalancing during the second half of the retrospective period in favor of investigations and away from studies. Evaluation studies can be, and some have been, very influential. But timely investigations—in view of their importance for institutional accountability—must trump studies in their command of resources, IEO Management attention, and scheduling priority. Given the scarcity of resources, the IEO will need to consider carefully how to prioritize investigation-oriented evaluations on institutional governance, financial crises, and/or implementation shortfalls in Board-approved policies as compared to more discretionary evaluation studies of programs and other activities.
- *IEO's unfinished business.* The IEO has usefully looked at IMF governance with respect to “Management and above” but there remains considerable scope for it to examine IMF governance from the perspective of “Management and below.” This retrospective review has highlighted critical questions about exactly how the IMF decides what position to take when institutional policies are not being fully implemented (as emerged in the IEO investigations of compliance with IMF operational policies) or when there is not an agreed policy (as emerged in the IEO studies of soft mandates). Pending a new evaluation on internal governance, ongoing evaluations could usefully focus on documenting cross-country differences in IMF treatment and probing their causes, as a basis for recommending possible remedies for institutional and/or staff practices.

### *IEO Work Program Execution*

- *IEO's data and methods.* The IEO has learned from experience in executing evaluations. Evaluation tools and data management—especially for surveys—have evolved somewhat as new IEO teams learned from earlier IEO work and innovations. And recent evaluations and evaluation completion reports have refined the IEO's approach to structured interviews. Nonetheless, there is room for greater efficiency in data management and for more systematic approaches across evaluations, though moves in this direction would need to be weighed against possible increases in implementation costs. Meanwhile, to support and complement ongoing IEO efforts to improve its handling of IMF documents, clearer guidance is needed to IEO teams on the use of these documents, especially with respect to two issues: (1) the consistent coding of cross-country documents, so the embodied evidence can be reliably quantified and analyzed—this is clearly an area that warrants a close watch going forward; and (2) the inclusion of quotations and paraphrasing in evaluation reports, while remaining within the strictures of existing protocols safeguarding confidential IMF material. With the IEO serving as the eyes and ears of Executive Directors and external stakeholders, effective and appropriate use of such material is paramount.

- *IEO's findings and recommendations.* There are two main takeaway messages here: First, the number and presentation of recommendations have varied widely across evaluations, with some recurring recommendations, but often without clear prioritization of recommendations and in some cases without clarity on their genesis. To date, the 18 evaluations have averaged six headline recommendations plus nine sub-recommendations each, with some evaluations containing double those amounts. Large numbers of recommendations invite treatment as menus rather than priorities and blur IEO's accountability. It will be helpful for the IEO to be clearer, more systematic, and consistently brief about what it is recommending as priority actions, what it is advising as possible actions, and what it is sharing as ideas for consideration. Second, IEO evaluation reports have not all made clear the logical framework underlying their progression from evidence-based findings to recommendations. Opinions can and do vary on how best to present this progression, in large part depending on views about who constitutes the target audience. For users who want a quick read-out of the IEO findings and conclusions, a brief and bundled presentation in the executive summary is fine. But for those who see the IEO's main value in terms of the evidence it is able to assemble, drawing on its privileged access to people, documents, and numbers, appropriate unbundling into the evaluation's framework of facts, diagnosis, and recommendations—perhaps presented in an annex—is essential.