

Methodology for Peer Reviews

This annex explains the methodology used in carrying out the peer review of five product lines of IMF research: the *World Economic Outlook (WEO)*, the *Global Financial Stability Report (GFSR)*, the *Regional Economic Outlooks (REOs)*, the selected issues papers (SIPs), and the working papers (WPs).³⁰ It describes the selection of external experts and the framework for the review, and it includes the rating templates. The findings and conclusions of these reviews are included in five background papers accompanying this report.³¹ The reviews assessed the quality of these product lines based on these experts' knowledge of the field, and considering the goals of each type of research product. Each background paper was prepared by an expert or group of experts on the corresponding issues. Experts were selected taking into account their experience in utilizing research and their familiarity with the different products, in addition to their substantive knowledge.

WEOs and GFSRs. All analytical chapters of the WEOs and the GFSRs for the period 2004–08 were assessed. The primary reviewer was a former deputy minister of finance from a large emerging market economy who also had academic and private sector experience. All chapters were reviewed independently a second time by either an academic or an investment banker.

REOs. For the *REOs*, all publicly available *REOs* from 2003 to 2009 were reviewed in their entirety, with an emphasis on the analytical chapters. The primary reviewer was an academic with considerable experience working in international financial institutions.

SIPs. A sample of 60 SIPs from 30 countries was reviewed. The universe of SIPs was divided into three income groups: low-income countries (LICs), 290

papers; middle-income countries (MICs), 472 papers; and high-income countries (HICs), 348 papers. Twelve countries were selected from each of the first two groups and six from the third group, making sure that the selection included countries covered by all area departments in the IMF and deliberately over-representing the LIC and MIC groups. After selecting the countries, a random number generator was used to select two SIPs per country. There were two primary reviewers for the SIPs. The first reviewer was a former chief economist at an international financial institution. The second reviewer was a former central bank governor, and currently an academic and investment banker. In addition, a secondary reviewer, an academic with policy experience, contributed to the assessment.

WPs. Two separate peer reviews were conducted for working papers: one for working papers on monetary policy frameworks, and the second on tax and revenue issues. For each theme, 60 working papers were randomly selected from the universe of papers issued during 1999–2008 (between a third and a half of all working papers on these themes). Two independent panels of three academics each were brought together to assess the technical quality of each sample of 60 papers. Each academic was considered an expert in his/her field (including being editors of lead field journals) and had some policymaking experience.

Peer reviews across all product lines followed a similar framework, embodied in the templates presented below. Templates follow some common core concepts, even if wording and syntax differ somewhat reflecting differences in the nature of each product. The four core assessment concepts are: (1) framework, (2) analysis, (3) output, and (4) policy conclusions or policy relevance. A five-point rating scale was used in each review with, again, some changes in the presentation. For example, for working papers the ratings were labeled “superior” (S), “above average” (AA), “average” (A), “below average” (BA), and “unsatisfactory” (U).

³⁰ Other product lines were reviewed by the evaluation team using a similar framework.

³¹ See IEO Background Papers BP/11/02 to BP/11/06 (Kuttner and others, 2011; Boadway and others, 2011; Selowsky and Škreb, 2011; Kiguel, 2011; and Montiel, 2011). Abstracts of these papers are in Annex 4 and the papers are available at www.ieo-imf.org.

A slightly modified labeling of rating scale was used for selected issues papers—among the differences, the rating of 3 was labeled “acceptable” rather than

“average.” The rating scale for the *WEO* and *GFSR* used numerical ratings only, ranging from 1 to 5, with 1 being the lowest and 5 being the highest.

World Economic Outlook/Global Financial Stability Report

Evaluation criterion	Rating ¹				
	5	4	3	2	1
Structure and presentation					
1. The questions are well posed and clearly focused					
2. There is a sound and clear discussion of the analytical framework					
3. There is a good balance between analytical discussions, empirical evidence and policy implications					
4. Writing is clear and well organized					
5. Is the empirical evidence presented in an intuitive and convincing way?					
Choice of topics					
6. The issue is timely and relevant					
7. The issues are important for policy discussions at the time					
Quality and analysis					
8. Does it use appropriate analytical tools and relevant data?					
Policy advice					
9. Would it be useful for policymakers?					
10. Are the policy issues convincing?					
Overall rating chapter					

¹ Kiguel (2011) used numerical ratings of 1–5, with 1 being the lowest and 5 being the highest.

Regional Economic Outlook

Evaluation criterion	Rating ¹				
	S	AA	A	BA	U
1. Adequacy of information to form an independent opinion					
2. Provision of a regional framework					
3. Clarity, nonambiguity, and thoroughness of content					
4. Grounding of analytics in the relevant literature					

¹ Montiel (2011) rated REOs by product, not by chapter, using a five-point rating scale.

Selected Issues Papers

Evaluation criterion	Rating ¹				
	E	VG	S	MU	U
Framework					
1. Question is clearly posed and its relevance to the country well articulated					
Analysis					
2. Uses an appropriate theoretical/conceptual framework					
3. Uses appropriate data and empirical methods proficiently					
4. Includes critical discussion and/or robustness analysis of results					
Output					
5. Writing is clear and well organized					
6. Conclusions are firmly grounded on the analysis					
7. Articulates clearly the policy relevance of findings for the country					
Overall rating					

¹ Selowsky and Škreb (2011) use a five-scale rating from 5 to 1, using the following labels E = "Excellent," VG = "Very Good," S = "Satisfactory," MU = "Moderately Unsatisfactory," and U = "Unsatisfactory."

Working Papers

Evaluation criterion	Rating ¹				
	S	AA	A	BA	U
Framework					
1. Question is well posed and clearly focused					
2. Places work within the context of existing literature					
3. Specifies contribution to existing literature					
Analysis					
4. Uses an appropriate theoretical/conceptual framework ²					
5. Uses appropriate data and empirical methods proficiently					
6. Includes critical discussion and/or robustness analysis of results					
Output					
7. Writing is clear and well organized					
8. Adds value relative to existing research					
9. Conclusions are firmly grounded on the analysis					
10. Articulates policy relevance of findings					
Overall rating					

¹ In the two background papers (Kuttner and others, 2011 and Boadway and others, 2011) on working papers, each reviewer assigned ratings.

² Includes whether there was excessive use of technique being used relative to the question being posed.