

the savings mainly driven by lower costs in advanced economies. Reflecting the latter factor, average costs for the IMF alone have fallen by about 10 percent.

- Since 2003, there has been surprisingly little difference between the average costs of FSAPs for advanced, emerging market, and low-income countries. This reflects the bigger impact of the 2003 streamlining measures on advanced economy FSAPs and the fact that some of the most complex financial systems were assessed prior to 2003.
- Average direct costs of FSAPs undertaken for low-income countries have not declined and are as expensive as those for emerging economies. About 60 percent of these costs are still borne by the IMF.

vi) **There is inadequate discussion of the expected scope of the FSAP, including with the authorities, at the terms of reference (TOR) stage.** Our reviews of the TOR for the 25-country sample found only a few cases where there was a serious initial discussion of priority-setting and strategic tradeoffs—although all of these cases were relatively recent, which suggests some improvement over time.¹⁸ Many country authorities said greater consultation at the TOR stage would have made the process more effective; some said they were surprised at a late stage by aspects of the FSAP’s scope for their countries. Our interviews with IMF staff suggest that the scope of FSAP Updates is a source of debate between Bank and Fund staff, with Bank staff frequently pressing (successfully) for a larger scope—to address medium-term development issues not taken up in the original FSAPs—whereas the IMF would have preferred smaller Updates focused on a follow-up on core issues from the earlier FSAP.

vii) **The assessment of the (AML/CFT) standard has little integration with other FSAP activities.** This view was broadly shared by country officials and FSAP team members. In practice, there was little synergy with other FSAP activities because of the special legal and accounting aspects involved, which required a different type of expertise.¹⁹

III. QUALITY OF FSAP PROCESSES AND DIAGNOSTIC TOOLS

12. We discuss here the efficiency of FSAP processes and views on the technical quality of the FSAP teams before going on to discuss various components of the FSAP output—the macroprudential analysis, the standards and codes assessments, and how effectively the various diagnostic elements are integrated into a comprehensive overall assessment with clear and well-prioritized recommendations.

¹⁸ Recent good practice examples include Chile (initial FSAP) and Ghana (Update).

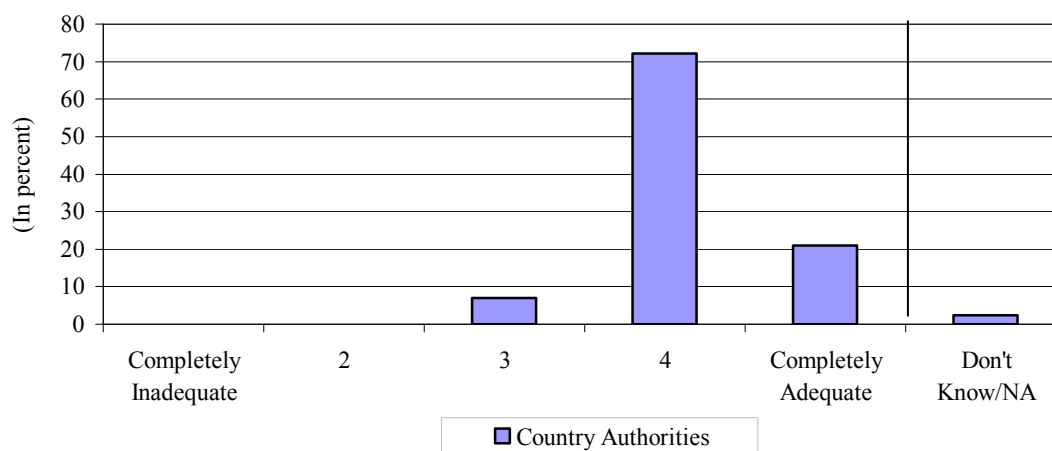
¹⁹ This says nothing about the effectiveness or value added of the AML/CFT assessments, which was not part of the evaluation’s terms of reference.

A. Efficiency of FSAP Inputs and Processes

13. Our in-depth reviews of 25 cases as well as interviews and surveys of officials and IMF and World Bank staff suggest the following main messages with regard to organizational aspects of the FSAP:

i) **Country authorities generally rated the technical quality of the FSAP teams highly, particularly the expertise of specialists.** Both our in-depth interviews with officials and the authorities' survey results (see Figure 2) suggest a high degree of satisfaction with FSAP teams' technical skills. A large proportion of officials we interviewed said that they viewed the opportunity to interact with the FSAP technical experts as a major value added from the exercise; indeed, many would have liked to have had more structured arrangements to follow up on specific issues with the experts concerned.

Figure 2. Assessment of FSAP Team's Technical Skills

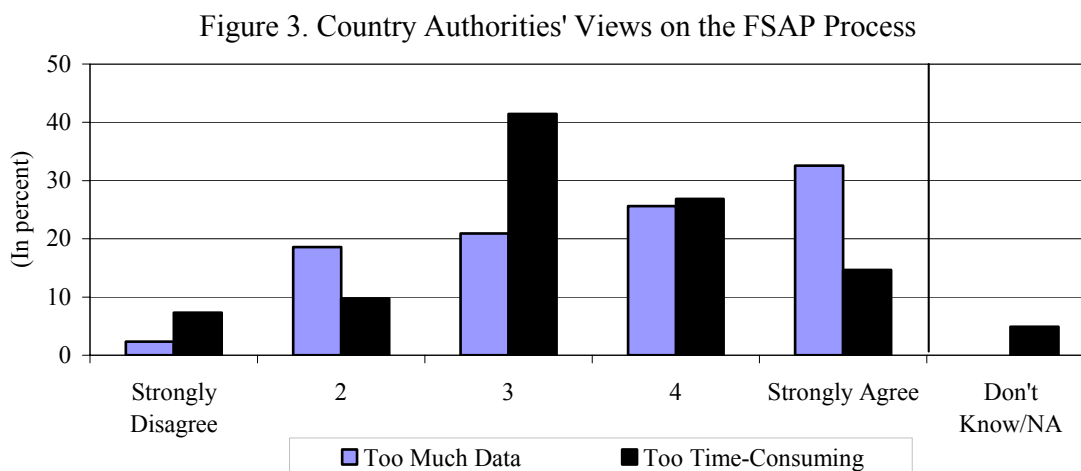


Source: Q6 of the survey of country authorities.

ii) **However, insufficient time for the FSAP team to prepare and familiarize with country-specific circumstances was a widespread complaint—noted by many authorities and, to a lesser extent, by the teams themselves.** In a number of cases, greater consultation with the authorities at an early stage of the process (i.e., the TOR stage) would have provided guidance on the most relevant expertise.²⁰ Interviews also suggest some shortcomings in the integration of the FSAP technical expertise with area department country-specific knowledge.

²⁰ For example, in a number of Eastern European countries looking toward membership of the European Union, cross-country experience on how a number of regulatory issues were dealt with in various EU members was of particular interest. However, some of the technical experts on the FSAP teams, while highly qualified, did not have the appropriate background to provide such information.

iii) **The burden of the FSAP on the authorities is inevitably very high, but could be eased somewhat by better planning.** While we have not been able to obtain any specific estimates of the costs the FSAP imposes on the authorities, the resource inputs required have strained the capacity of even well-trained and well-funded supervisory systems, especially when extensive translation of documents into English was required.²¹ A large proportion of survey respondents were of the view that the time and data requirements of the exercise were excessive (Figure 3). The in-depth examination of the 25 country cases indicated that a large part of these costs were intrinsic to the exercise, and many officials recognized that the extremely intensive data gathering had eventually yielded benefits in terms of better data for macro-prudential analysis or greater transparency about a country's financial system and regulatory approaches.²² Nevertheless, the burden could be significantly eased by (a) better planning and consultation at an early stage to take account of country circumstances, leading to greater selectivity in information requests; (b) greater lead time on questionnaires and data requests; (c) greater personnel continuity from previous financial sector work and in any follow-up work; and (d) possibly preparing some of the ROSCs in advance of the main FSAP mission.



Source: Q7.1 and 7.3 of the survey of country authorities.

²¹ For example, in the case of Japan senior officials noted that at some stage in the process about 10 percent of staff of the Financial Services Agency were involved in the FSAP exercise—at a time when there was strong need to attend to the financial sector's difficulties.

²² For example, in Korea, senior officials expressed the view that the FSAP had very high data requirements (e.g., on cross-sectoral data for stress testing analysis of the corporate sector), but these subsequently proved very useful for their continued assessment of the financial sector and exposures to the corporate sector.

iv) **The choice of FSAP team leaders is critical.** The team leader (and deputy) play a crucial role in identifying priorities for the assessment and integrating the results of the various diagnostic instruments into an overall assessment with clear, well-prioritized recommendations. This is an enormously challenging task for which considerable technical expertise and policy judgment is required. Our interviews with some IMF and World Bank senior staff indicate a concern that, as the FSAP becomes more “routine”, less-experienced team leaders have begun to be chosen, with a potentially adverse impact on quality (e.g., a tendency to follow a “template” approach without a deep understanding of the situation in each country).

v) **From the IMF perspective, the preparation of separate documents (the FSSA and the FSA, respectively) for the Board of the IMF and World Bank appears to have helped minimize delays and the burden of tailoring the FSAP results to different institutional needs.** However, the FSSA is better anchored in IMF processes (Article IV) than the FSA is in those of the Bank (see OED report) and FSSA reports are produced with a significantly shorter time lag than the FSAs.²³ A number of IMF staff emphasized that anchoring the FSSA in the Article IV surveillance process had imposed a clear timetable that had helped to avoid excessively drawn out discussions—both with the authorities and among the FSAP team—on details.

B. Macro-Prudential Risk Analysis

14. A message from our interviews, reinforced by the survey results, is that **in many, but not all, countries the FSAP has contributed significantly to assessing financial sector vulnerabilities—by helping to change the culture towards one that emphasizes system-wide risk assessments and, in many cases, upgrading methodologies. Within this overall positive experience, however, there are significant differences across countries, and several shortcomings need to be addressed.**

15. The two main diagnostic tools used in FSAPs for analyzing macro-prudential risks of financial systems are, first, stress-testing how different measures of financial strength (e.g., capital adequacy and profitability) would respond to a variety of shocks and, second, analyzing trends in various financial soundness indicators (FSIs). The principal conclusions are as follows (see Annex V for more details):

i) **The use of methodologies for stress-testing at the level of the overall financial sector is still in its infancy.** The degree of sophistication of approaches used varies substantially across FSAPs, depending in large part on data availability, cooperation with the authorities, time available for the analysis and the judgment of the FSAP team (see Box 1 for

²³ The average time lag between the date of the first FSAP mission and circulation of the FSSA is just under 40 weeks; for FSAs, the average time lag is about 58 weeks; no FSAs for FSAP Updates had been completed as of end-August 2005.

some good practice characteristics encountered in the country reviews). But even with relatively “sophisticated” approaches, the results obtained can depend critically on how various shocks are calibrated and feedback effects modeled. In practice, data and other limitations constrained the use of stress-testing to fairly basic approaches. For example, in almost half of the 25 cases examined in depth, the principal methodology for analyzing credit risk of the banking sector was based on a simple static exercise that assumed (relatively arbitrary) increases in levels of banks’ nonperforming loans together with assumptions on different provisioning levels. Even rudimentary tests can add value, especially when undertaken in conjunction with other analysis, but the limitations of such approaches need to be clearly flagged.

Box 1. Good Practices on Stress Testing

Stress testing is a method for quantifying the impact of future extreme but plausible shocks on a financial system. The degree of sophistication of approaches used varies substantially across FSAPs, depending in large part on data availability, sophistication of the financial system, cooperation from the authorities, time available for the analysis, and the judgment of the FSAP team. We summarize here a number of “good practice” approaches to different aspects of such tests, drawn from the 25-country example (see Annex V for further details). 1/

Data quality. The quality of data, and its implications, for any results should be described candidly; many FSAPs are weak in this respect; (Cameroon is a good practice exception). There are some cases where the available data is of poor quality and where vulnerabilities are fairly obvious. Not conducting stress testing should always be an alternative in such cases, as otherwise there is a high risk of spuriously concrete results that mask an unknown situation (e.g., the Costa Rica FSAP appropriately did not undertake any formal stress tests).

Scenarios and events. Most stress tests have included single factor sensitivity analysis. The most recent vintages (e.g., Jordan and New Zealand and many European countries) have also included the use of scenarios that involve simultaneous movement in various macro risk factors. This is a positive trend, as such scenarios could help analyze better the vulnerabilities of the financial system

Calibration of shocks. The challenge is to be able to have a common understanding for what can be considered exceptional but plausible shocks. Where feasible, the calibration techniques could use models to characterize the relationships among macro risk factors in the context of different scenarios and/or cases in which single variables are shocked (by using statistical or historical approaches). For example, some recent FSAPs (Germany and Chile) have derived a consistent set of shocks to macro variables from a macro model.

Methodologies. While it is often necessary to tailor an FSAP stress test to data availability and the sophistication of the financial system, it would be useful to form “country peer groups” based on some criteria related to the complexity and sophistication of a financial system. Standardizing a core set of data sets, methodologies and sensitivity analysis within the peer group could lead to the development of common benchmarks for cross country comparisons, thus facilitating vulnerability analyses. For example, for the group of industrialized countries, stress testing should aim to move towards a common good practices set of methodologies.

Interpretation of the results. More attention needs to be given to the interpretation of stress test results, not only in light of the methodological caveats but also in terms of the relative importance of different shocks (e.g. avoid overemphasizing market risks when credit risks are more relevant from a vulnerability perspective). This is an area where many FSAPs are weak, but Korea and Cameroon are good practice examples.

1/ Mention of a country’s FSAP with regard to one aspect does not necessarily mean it was “good practice” in other respects.

- ii) **The reporting of results in most FSAPs tends to downplay these limitations and often reports the bottom line results from stress-testing as if from a “black box” exercise.** This often results in overly simplistic messages about the strength of the financial sector. Greater “health warnings” about the interpretation of results are needed, especially when the quality of the data is weak.
- iii) **There is a considerable gap between the “good practice” approaches to modeling shocks and those used in many other cases.** A number of recent FSAPs (including many for European countries, but also Chile) have generated a consistent series of shocks to specific macro variables derived from a macro model. In other cases, a series of shocks to particular variables have been aggregated but without being derived from a clearly defined, consistent macroeconomic model.²⁴ Stress tests are supposed to analyze “exceptional but plausible” scenarios of shocks, but in only about half of the cases we examined was there an attempt to provide a clear rationale for the size and composition of the shocks chosen.
- iv) **Some assessments have avoided analyzing the consequences of politically sensitive shocks (e.g., public debt defaults).** While there is an understandable tendency not to rock the boat by focusing on such major potential adverse events, the result could be reassuring statements in the FSAP that the financial system is robust to a variety of milder shocks, leaving it to each observer to read between lines with regard to the larger shocks. This could lead to potentially misleading signals. One possible alternative could be to adopt an approach where certain types of shock are considered in certain situations—for example, when the consequences for bank balance sheets of specific downgrades in sovereign public debt are to be analyzed—but without creating uniform sets of shocks that preclude adaptation to particular country circumstances.²⁵
- v) **There is still insufficient attention in many FSAPs to global and regional linkages, including for countries with substantial international capital market links.** The evaluation’s rating on incorporation of global and regional risks shows that consideration of these risks for the 25 country cases has fallen short of good practice in a significant

²⁴ In the case of Japan, problems associated with weak domestic “ownership” of the exercise also limited its usefulness. Agreement could not be reached on the nature of the tests to be undertaken and the authorities eventually asked that the data they had provided not be used. The FSAP team undertook at a late stage a number of tests using more limited data available from published balance sheets of banks.

²⁵ For example, one approach could be to include in the group of shocks “tested” the impact on interest risk premiums from a downgrade in public debt by a pre-specified number of grades whenever the initial credit rating is below a particular threshold. This suggestion would be consistent with the Basel II capital standards methodology which proposes a system of risk weightings (at least for foreign-currency denominated debt) based on the sovereign’s credit rating.

proportion of cases (e.g., about one-third of cases were assessed as having some problems in this respect; see Table 2). Moreover, the evaluation's average internal rating in this respect was even lower (2.2 on the 4-point scale) for those countries judged to be of global or regional systemic importance. The evaluation found generally little analysis of cross-border linkages capable of spillover effects even in some countries with global systemic importance (e.g., Japan and Russia). Also, the Ireland and Singapore FSAPs focus their analysis on the vulnerabilities of local banks or the foreign banks that have domestic operations. They make reference to the linkages of domestic banks with the international financial centers but do not analyze the risks of these centers in any detail, as it is assumed they do not belong to any country in particular.²⁶

Table 2. Results of the IEO Assessments of FSAP Content 1/

Criteria	Mean Score (on scale of 1–4)	Percentage of ratings indicating some problems (i.e., ratings of 3 or 4)
Extent of incorporation of regional and global risks into analysis	1.84	28
Balance of development and stability issues	1.88	16
Integration of standards and codes in overall assessment	1.84	20
Coverage of overall financial sector 2/	2.00	20
Clarity and candor of findings	1.88	12
Importance and consequence well explained	1.94	20
Clarity of recommendations	1.82	8
Usability of recommendations (e.g., specificity)	1.96	16
Prioritization of recommendations	2.46	44
Degree of alignment of FSAP and FSSA 3/	1.42	0

1/ IEO assessors rated each of the FSAPs for 25 countries in depth with respect to the above criteria. Each aspect was rated on a four-point scale (with 1 being the highest). To minimize subjective judgments, the evaluations were guided by a detailed template of what would be expected to achieve specific ratings for each category (see Annex III).

2/ This refers to structure, trends, and links to other sectors.

3/ Refers to extent to which key messages are candidly reflected in both publications.

vi) **FSIs have generally not yet been used in a meaningful manner in most assessments (reflecting problems with data and interpretation of appropriate benchmarks for signaling vulnerability as well as inadequate time series).** Although most FSAP reports include tables on FSIs, in only half of the 25 cases examined in depth did the reports provide some interpretation in terms of the risk implications of the figures. Since

²⁶ One “good practice” exception where cross-border links were explicitly incorporated is the New Zealand FSAP; in light of the substantial role of Australian-owned banks in New Zealand, the mission discussed the performance of these banks with Australian regulatory authorities.

sector-wide averages may mask concerns with specific groups, for analyzing potential vulnerabilities aggregate indicators frequently need to be complemented by indicators for key peer groups within the banking sector (i.e., state banks, foreign banks, local private banks). More generally, interviews with area department staff indicated that many felt they lacked the necessary training and experience to interpret FSIs and integrate the analysis into ongoing surveillance work, even when the data were available on a consistent basis.

vii) **The quality of the data on the financial system is often not emphasized sufficiently.** In some countries, more caution is needed before using available statistical data at face value, either for stress-testing or other analysis.

C. Standards and Codes Assessments

16. The evaluation reviewed only assessments of standards and codes prepared as part of FSAP exercises.²⁷ Drawing on the 25 in-depth case reviews, interviews with country officials, and IMF and World Bank staff as well as the survey results, the principal conclusions are as follows:

- i) **There is no evidence that the streamlining of the number of standards assessed in detail has created problems for the ability of FSAPs to make overall judgments on financial sector vulnerabilities, but the rationale for which standards to assess is not discussed sufficiently.** A “scoping exercise” is first conducted to identify the set of standards to be assessed in each FSAP, but in the cases we reviewed there was often limited discussion in the TOR or the subsequent FSAP reports of why choices are made (e.g., why insurance or securities standards were covered in some low-income countries where the sector was very small in relation to GDP or why payments system standards were or were not covered). However, the evaluation did not identify any cases where omission of a detailed standards assessment had contributed to significant shortcomings in the overall assessment of potential vulnerabilities. The review of all post-2003 FSAP cases confirmed these conclusions.
- ii) **While the assessments generally distinguish between de jure standards and de facto implementation, the significance of institutional weaknesses is often not emphasized sufficiently.** In most of the 25-country sample, assessors did take account of

²⁷ See Appendix Table 3 for a full list of the various standards and codes. Those most often covered in FSAPs are the Basel Committee’s Core Principles for Effective Banking Supervision (BCP), International Association of Insurance Supervisors (IAIS), International Organization of Securities Commissions (IOSCO), and Committee on Payments and Settlement Systems (CPSS), which were the main focus of the evaluation; the banking supervision standards were assessed in all FSAPs. For a discussion of other standards and codes, see *The Standards and Codes Initiative—Is it Effective? And How Can it Be Improved?* (SM/05/252, July 1, 2005). As noted earlier, assessment of the AML/CFT standard was not part of the terms of reference of this evaluation.

differences between de jure laws and regulations and de facto implementation. Indeed, the assessment methodologies to some degree require making this distinction—i.e., they require interpretation of compliance and evidence of enforcement or non-enforcement of various principles.²⁸ However, while many FSAP reports do discuss problems in forbearance in regulations or low enforcement capabilities, it is sometimes difficult to read between the lines to judge the severity of the potential macroeconomic significance of such shortcomings (e.g., Dominican Republic, Sri Lanka, Tunisia). One “good practice” example where enforcement issues are explicitly linked to the vulnerability analysis is the Ghana FSAP.

iii) **Integration of the various standards assessments into an overall FSAP assessment does seem to have added value, but the degree of integration varied from case to case.** The review of the 25 country cases suggests that incorporation of the standards assessments into a broader discussion of financial sector vulnerability and development issues did add value in many cases, especially for the banking sector standards.²⁹ In some cases, however, a “headcount” approach to listing performance vis-à-vis various principles was not accompanied by a sufficiently integrated discussion of the potential impact of various identified shortcomings (e.g., in the Egypt, Philippines, and Romania FSAPs).

iv) **A number of officials noted that an excessive focus on the “number” of principles for which a country was fully or largely compliant could give a misleading signal on the potential downside consequences of remaining gaps.**³⁰ Interviews with staff and authorities indicate that there were often greater disagreements on the ratings than on the underlying qualitative assessment. In its recent review of the Standards and Codes Initiative, the IMF Executive Board endorsed a number of changes to the presentation of ROSC findings, including a principle-by-principle summary of the observance of each standard and an executive summary providing a clear assessment of the overall degree of observance of

²⁸ See, for example, the Basel Core Principles methodology (available at www.bis.org).

²⁹ This conclusion is supported by the survey of IMF surveillance mission chiefs conducted as part of the internal review of the Standards and Codes Initiative. A high proportion of mission chiefs—especially for emerging market and developing countries—were of the view that the standards assessments had added value to the usefulness of the FSAP for surveillance. See *The Standards and Codes Initiative*, op cit, Background Paper (SM/05/252, Supplement 1), Table 15. See also the assessment of the overall FSAP content in the next section below.

³⁰ To give only one example, senior Egyptian officials noted that their efforts to strengthen further the supervisory framework initially encountered an internal reaction from many supervisory staff that the high proportion of “compliant” ratings under the BCP assessment meant that there was no need for substantial additional efforts.

the standard, while avoiding a rating or “pass or fail” report.³¹ While it is too early to judge the effect of such changes, this evaluation reinforces the view that the overall qualitative assessment and identification of key remaining gaps are the most critical elements and that the exercise should not be condensed solely to one of ratings, even if that is the aspect that market participants indicate that they value the most (see below).

v) **The governance structure for assessing standards is a little vague, but present arrangements for providing feedback work satisfactorily in practice.** The issue is—“who assesses the assessors?” In principle, this is the responsibility of the IMF and World Bank Boards for those standards assessed under the FSAP. In practice, members of the Board are not in a position (e.g., they are not provided with the necessary information) to make such judgments. The various standard-setting bodies, and their secretariats, do have the appropriate background but do not have governance responsibility for assessing whether the assessment exercises are proceeding satisfactorily; indeed, they do not even see those FSSAs that countries do not agree to publish. Nevertheless, discussions with the various standard-setting groups suggest that, in practice, there are sufficient informal and formal channels (including the Financial Sector Forum and IMF/World Bank staff participation in various technical committees) for adequate feedback to be provided on how assessments are being conducted. Our interviews with the various secretariats suggests a high degree of satisfaction with the results (see below).

IV. FSAP CONTENT

17. In judging the overall quality of the FSAP content, we relied upon two major sources of evidence. First, IEO assessors rated the content of the FSAPs for the 25 country cases on a 4-point scale according to various criteria: coverage and balance of assessments; clarity and candor of findings; as well as clarity, usability and prioritization of recommendations (see Table 2). These ratings on individual components were also used as inputs into an overall qualitative judgment on how well the FSAP assessment was integrated across the various sectors and with the overall macroeconomic picture. A high quality overall assessment is one that combines effectively the results from the various evaluation tools to present the main risks and vulnerabilities to the financial sector—i.e., those of macroeconomic/systemic significance—with an indication of criticality and consequence. Second, the surveys of various groups of stakeholders included questions on various aspects of the quality of FSAPs. We discuss first the overall quality of the FSAP assessment and then the articulation of findings and recommendations. Finally, some issues related to the “joint” IMF-World Bank nature of the FSAP are addressed.

³¹ See *The Acting Chair’s Summing Up: The Standards and Codes Initiative—Is it Effective? And How Can it be Improved?* July 25, 2005.