

QUALITY OF ESW IN FY02

A QAG ASSESSMENT

APRIL 1, 2003



ABBREVIATIONS

AAA	Analytic and Advisory Activities	IDA	International Development Association
ACS	Activity Completion Summary	IMF	International Monetary Fund
AFR	Africa Region	JSA	Joint Staff Assessment
ARPP	Annual Review of Portfolio Performance	LCR	Latin America and the Caribbean Region
BB	Bank Budget	MNA	Middle East and North Africa Region
CAS	Country Assistance Strategy	OA	Overall Quality Assessment
CEM	Country Economic Memorandum	OED	Operations Evaluations Department
CFAA	Country Financial Accountability Assessment	PA	Poverty Assessment
CPAR	Country Procurement Assessment Report	PER	Public Expenditure Review
CPIA	Country Policy and Institutional Assessment	PREM	Poverty Reduction and Economic Management Network
D&D	Dialogue and Dissemination	PRSC	Poverty Reduction Support Credit
EAP	East Asia and Pacific Region	PRSP	Poverty Reduction Strategy Paper
ECA	Europe and Central Asia	PSI	Private Sector Development and Infrastructure Network
ESSD	Environmentally and Socially Sustainable Development Network	QAG	Quality Assurance Group
ESW	Economic and Sector Work	QER	Quality Enhancement Review
FSA	Financial Sector Assessment	SAR	South Asia Region
FSAP	Financial Sector Assessment Program	SAP	Systems Applications Program
FSSA	Financial System Stability Assessment	TA	Technical Assistance
FSE	Financial Sector Network	TF	Trust Fund
FSN	Financial Sector Note	TTL	Task Team Leader
FY	Fiscal Year	WDR	World Development Report
HDN	Human Development Network		
HQ	Headquarters (i.e. Washington, DC)		
IBRD	International Bank for Reconstruction and Development		

ACKNOWLEDGEMENT

This report was prepared by a team led by Praveen Kumar and Kadir T. Yurukoglu, and included Robert Drysdale and Bing Deng. Dora Adoteye, Leila Cruz, Susan Crisostomo, Vinod Ghosh, Ann Walters and Lora Marshall provided administrative and logistical support. A wealth of ideas came from both our panelists and the task teams selected for the QAG sample.

Prem C. Garg, Director, Quality Assurance Group, guided the overall effort.

**QUALITY OF ESW IN FY02
A QAG ASSESSMENT
TABLE OF CONTENTS**

EXECUTIVE SUMMARY	i
I. CONTEXT	1
II. OBJECTIVES AND APPROACH	4
III. ASSESSMENT FINDINGS	8
A. Bank-wide Results	8
B. Task Development Objectives.....	11
C. Task Context.....	13
D. Results by Quality Dimension	13
E. Results by Region	19
F. Results by Networks and Products.....	21
IV. BANK INPUTS AND PROCESSES	26
V. PILOT ASSESSMENT FINDINGS	29
A. Country ESW Assessment	29
B. Financial Sector Assessment Program	32
C. Other ESW	33
VI. RETROSPECTIVE AND NEXT STEPS	37
A. Retrospective	37
B. Next Steps	38
C. Feedback on the Assessment Process	38
VII. CONCLUSIONS AND RECOMMENDATIONS	40
 ANNEXES	
Annex 1 Main Contributors	41
Annex 2 List of Sampled ESW Tasks.....	42
Annex 3 FY02 Approach Paper	43
Annex 4 Detailed Quantitative Results	
□ Table 1 - Results by Region	51
□ Table 2 - Results by Network.....	52
□ Table 3 - Results by Product Type	53
□ Table 4 - Results by Cost Range.....	54
□ Table 5 - Results by HQ/Field Managed	54
□ Table 6 - Results by IDA Status.....	55
□ Table 7 - Results by CPIA Category.....	56
□ Table 8 - Results by Participation Level.....	57
□ Table 9 - Average Elapsed Time and Costs by Group	58

Annex 5	Guidance Questionnaire with Bank-wide Percentage	59
Annex 6	TTL and Panelists Surveys	
	□ Table 1- TTL Survey	66
	□ Table 2- Panelist Survey	67

TEXT TABLES

Table 3.1	A Pattern of Excellence	9
Table 3.2	Task Development Objectives	12
Table 3.3	Strengths and Weaknesses by Region	20
Table 3.4	Strengths and Weaknesses by Network	22

TEXT FIGURES

Figure 3.1	Overall Results.....	8
Figure 3.2	ESW Quality by IBRD/IDA Status	10
Figure 3.3	Results by Quality Dimension	14
Figure 3.4	ESW Quality by Participation Level.....	19
Figure 3.5	Results by Region	20
Figure 3.6	Results by Network.....	21
Figure 5.1	Other ESW by Quality Dimension	35

TEXT BOXES

Box 2.1	ESW in the FY02 ARPP	2
Box 2.2	FY02 ESW Sample	4
Box 3.1	Client/Partner/Stakeholder Feedback to Panelists.....	15
Box 3.2	Weaknesses in Attention to Poverty Reduction in ESW Tasks	17
Box 3.3	Weaknesses in Attention to Gender in ESW Tasks.....	17
Box 6.1	Scorecard on FY01 ESW Recommendations	37

EXECUTIVE SUMMARY

Economic and Sector Work (ESW) has acquired a major role in Bank operations. In the context of the CDF/PRSP as well as the ‘knowledge’ Bank, ESW has grown in FY02 to absorb about \$86 million of Bank budget—approximately 75% of the resources devoted to preparing new lending. The quality of ESW is rightly seen as central to Bank operations.

As proposed in the ESW report for FY01, an Advisory Panel was constituted to review the ESW assessment approach, take stock of progress and lessons, and chart the course forward. Its findings, endorsed by senior management, contained three main proposals: to carry out one further round of task-by-task assessments to verify if recent ESW quality gains are being sustained; to develop an ‘integrated’ approach to ESW assessment at the country or sector/thematic level; and, in accord with the new definition of ESW, to extend coverage of ESW assessments to other products—notably process tasks, joint products with the IMF, Network ESW and so on.

This fifth review of the ESW (ESW02) reports on the first stage of that agenda. It contains findings of a task-by-task assessment based on a random sample of 75 out of 228 ‘eligible’ tasks delivered to clients in FY02 (33% by number and 43% by cost). It also summarizes experience to date with three separate pilots: an integrated multi-year assessment of ESW-AAA in three countries; an assessment of the Financial Sector Assessment Program (FSAP) in four countries; and an evaluation of ‘Other ESW’-- informal tasks, such as conferences/workshops, country dialogue and consultations, of significant resource use, but not previously included in the ESW sample.

MAJOR FINDINGS

- **Overall Quality**¹ of ESW, based on the FY02 sample, was rated 94% *Satisfactory or better*, against a target of 90+%. This continues the improving trend. The *Highly Satisfactory* share is an unprecedented 22% by dollar cost. There were no *Unsatisfactory* tasks.
- Noteworthy gains in FY02 were a sharp increase in the *Likely Impact* of ESW tasks to 90% *Satisfactory* and a significant improvement in *Bank Inputs and Processes* to 83% *Satisfactory*, a reflection of more consistent and effective management attention to ESW quality.
- Performance has converged across all four dimensions (*Strategic Relevance, Internal Quality, Dialogue and Dissemination, and Likely Impact*) at around the 90% *Satisfactory* level.
- Both **Regionally and by Network**, quality has been sustained or enhanced, though some aspects merit closer attention. AFR, EAP, ECA, MNA and SAR have held or deepened earlier gains. LCR follows with room for improvement in both *Dialogue and Dissemination* and *Likely Impact*. Both EAP and LCR reveal uneven performance in the efficiency of resource use and management aspects, as gauged by

¹ Unless otherwise stated, ratings on ESW quality throughout this report are on a **dollar cost** basis. The corresponding percentages represent the proportion rated *Satisfactory (or better)* i.e. *Satisfactory* plus *Highly Satisfactory*. The ratings for *Bank Inputs and Processes* are on the basis of the **number** of tasks.

Bank Inputs and Processes. PSI Network shows much stronger performance across the board, though both PSI and PREM have room for further improvement in *Bank Inputs and Processes*.

- The **peer review** process is still not providing effective quality assurance. This applies throughout task implementation, including the decision draft, in contrast to evidence of more rigorous peer review standards at task inception.
- **Participatory ESW** continues to show important strengths, confirming the cumulative benefits of sustained engagement with a range of stakeholders throughout task implementation, notably the increased prospects of substantive impact.
- The average quality of **Core Diagnostic ESW**, a growing share of country ESW portfolio, has remained high. Consistent with that overall picture, the quality of CFAA/CPARs, which were drawn into the ESW sample for the first time, has improved well beyond the level identified in the FY01 pilot assessment. A special review of Public Expenditure Reviews (PERs) and of Poverty Assessments (PAs) in the sample reveals some unevenness in performance, particularly in aspects of task scope and the formulation of findings, conclusions and recommendations.
- With respect to sensitivity to **poverty focus and gender** in ESW, teams are developing more rigor in their approach to these cross-cutting themes, but, in many cases, they fail to exploit opportunities for poverty and gender analysis that fall within the task scope.
- **Success indicators.** Many tasks are launched without clear understanding as to what would constitute success. Interim benchmarks should be identified by TTLs as part of initial task (or program) planning. This conclusion emerges from a pilot initiative on the results orientation using task development objectives (DOs) and corresponding success indicators defined by task teams which were then assessed for their likelihood of being achieved.
- The use of **SAP** and the **Activity Completion Summary (ACS)** for task reporting and evaluation is highly problematic. Task teams continue to disregard SAP reporting. The relatively weak results orientation in ESW task management is reflected in the scattered and incomplete use of the ACS to record task evaluation --- including self-evaluation and client feedback.
- **Pilot Country AAA assessments** in three countries demonstrated the new insights that flow from assessing clusters of AAA within a country strategy context (CAS/PRSP). The approach captured important new lessons on program coherence, selectivity and task synergies.
- A pilot review of four **FSAPs** identified consistently high technical quality, but raised questions about the balance between ‘stability’ in the financial sense and broader ‘development’ objectives, the integration of FSAPs with Bank country strategy development and programming. There is a need to revisit FSAP processing guidelines to streamline requirements.
- A pilot assessment of 11 **“Other ESW Tasks”** (conferences/workshops, country dialogue, consultations) found the overall quality *Satisfactory* for 80% of these tasks. However, the assessment of *Bank Inputs and Processes* revealed weaknesses with respect to management oversight and reporting processes.

CONCLUSIONS AND RECOMMENDATIONS

Viewed against the stock of findings accumulated over five years of assessments, recent quality gains in ESW appear robust. Notwithstanding this impressive achievement, the FY02 assessment identified a few areas for further attention. The following actions are proposed:

For Regions and Networks:

- Regions, notably EAP and LCR, and the PSI and PREM Networks, should continue to focus on their quality assurance processes for ESW, including management oversight during task implementation/dissemination. They should also provide clear guidance on the peer review system as it applies, in particular, to the final stages of task implementation, notably the decision draft.
- PREM should make use of the detailed task assessments and summaries of the PERs and PAs in the FY02 sample to provide further feedback to the respective sector boards.
- Regions and Networks should continue to foster participatory approaches to ESW task design and implementation. They should highlight good practice to staff.

For OPCS:

- OPCS should encourage Regions and Networks/Sector Boards to improve guidance on incorporating task DOs and success indicators, including interim benchmarks, into ESW task (or program) planning.
- In consultation with Regions and Networks, OPCS should arrange to carry out a review of the quality of ESW information in the SAP and the prevalence of use and quality of the ACS. The purpose would be either to provide workable incentives for use of these tools as designed or to agree on alternative approaches to monitoring and reporting on ESW task progress and results.

For QAG:

- In partnership with Regions and Networks, QAG should move ahead with the next stage of piloting Country AAA assessment approach. An Approach Paper should be presented to Management for mainstreaming in FY04, drawing upon lessons learned from the pilots.

I. CONTEXT

1. Economic and Sector Work (ESW) provides the basis for the Bank's policy dialogue with clients, the development of country strategies, and the formulation and implementation of effective lending programs. ESW is also an important instrument for building institutional capacity, especially when it is undertaken with local partners. Like the Bank's other major business line, lending, ESW has undergone significant changes in recent years. It has become more client-focused and participatory. Starting with a modest sample in FY98, ESW has been subject to a task-by-task random assessment for five years, including this ESW02 exercise.

2. A new ESW governance framework was approved in 1998 covering themes such as strategic selection, cost-effectiveness and timeliness, and introduced the Activity Completion Summary (ACS) as a monitoring/reporting system. A reform program has been put in place, notably via the two reports on *Fixing ESW* in 2000 and 2001². In the first phase of this program (FY99-01), significant progress was made in cleaning up the pipeline of ESW reports, with enhanced quality assurance processes being introduced across the Regions. The second phase of the ESW reform (FY02-04) is focusing on improved coverage of Core Diagnostic ESW in country programs, raising the quality of ESW for IDA countries, better dissemination of lessons learnt from high-quality ESW, undertaking ESW in partnership with client countries to enhance country ownership and build local capacity, and pooling knowledge and harmonizing approaches with other donors.

3. In recognition of the greater diversity of ESW products and related activities, and as part of the effort to strengthen the governance of ESW, in January 2002 OPCS issued new instructions on ESW definition, classification and recording. ESW is now defined as an activity that involves analytic effort, is undertaken with the purpose of influencing an external client's policies, and is owned by a Bank unit. Under this more comprehensive definition, ESW is classified into five different categories of output: country diagnostic reports; country advisory reports; regional reports; other products (policy notes, conferences and other events); and process tasks (country analytic work that does not generate separate products). Bank units are requested to record tasks according to this classification.

4. Also in accordance with the higher priority assigned ESW within the Bank's operational programs, since FY01 the status of ESW has been reported within the Annual Review of Portfolio Performance (ARPP). The main findings of the FY02 ARPP pertaining to ESW are summarized in Box 2.1.

² *Fixing ESW: Where We Are*, OPS, 2000; and *Fixing ESW: Phase II, Challenges, and Next Steps in the ESW Reform Process*, OPCS, 2001.

BOX 2.1 ESW IN THE FY02 ARPP

The FY02 ARPP provides the following overview of ESW; as well as a summary of the ESW02 results:

- The role of ESW in total operational activities grew substantially in FY02, after several years of very limited growth. In direct Bank budget terms, in FY02 ESW absorbed \$86 million (FY01 \$48 million), compared with \$114 million for new lending.
- In terms of overall numbers, Regional ESW product deliveries in FY02 totaled 423, compared to 324 in FY01, an increase of 30 percent. Within this total, delivery of regular Reports was essentially unchanged between the two fiscal years. Rather, the growth was concentrated in a sub-category of Other Products, 189 tasks in number, which almost doubled in size in FY02.
- The program of Core Diagnostic ESW is now growing in line with the recommended five year rule for coverage. At end-FY02, OPCS estimated that coverage was about 40-50 percent, with the best coverage for PERs and the worst for CFAAs.
- A new SAP category has been created for Network-managed ESW. These products are managed and financed directly from the Sector Board anchor resources, and are quite substantial. These tasks numbered 109 in FY02 rising in FY03 to about 40 percent of Regional ESW activity. This activity should now be brought under the same regime as regular ESW for monitoring and quality assurance purposes.
- Heightened attention to ESW management has helped to strengthen the strategic focus of ESW tasks within the CAS program. Likewise, a more intense focus on a sound analytic basis for policy dialogue and project interventions has reinforced the commitment to task internal quality. In keeping with CDF/partnership principles, there has been a significant increase in the use of participatory approaches to ESW, with commensurate quality gains.
- Greater attention to defining ESW success indicators, and to monitoring and reporting on outcomes/early impact, is key to a stronger role for ESW within the CAS process and a more effective contribution to the *Results* agenda.

STATUS OF ASSESSMENT OF ESW

5. The Synthesis Report for the fourth assessment of ESW (the FY01 sample) proposed a “comprehensive review of the present ESW assessment approach ahead of the next cycle”³. That proposal was based on several considerations: the changing context of ESW noted above; cumulating evidence of progress in the quality of ESW based on the annual task-based assessments; as well as feedback on the assessment approach and methodology used over the four years of review. Accordingly, in March 2002, QAG commissioned an Advisory Panel on Assessing ESW. Under the leadership of David de Ferranti, that panel carried out a careful review of the experience with and lessons of ESW assessments, and also recommended an agenda for future ESW quality assessments⁴.

6. The main issues and recommendations reached by the Advisory Panel were as follows:

- a more integrated approach (for example, country ESW or sector/thematic ESW

³ *Quality of ESW in FY01*, p. 27, para. 64.

⁴ *Report Of The Advisory Panel On Assessing Economic And Sector Work*, May 7, 2002. See this report for the full range of findings and recommendations reached by the panel, of which only the main aspects are presented here and in Annex 3, para. 3

assessment) should be developed and tested

- the task-by-task approach should not include tasks under \$50,000, though they would be considered under the integrated approach
- the assessment of *Likely Impact* should be strengthened to the degree feasible by client feedback and also by clearer specification by task teams of the task development objectives (DOs) and the corresponding success indicators

7. With respect to the future direction of ESW assessments, the Advisory Panel reached the following conclusions which were endorsed by Bank senior management.

- FY03 would see:
 - continuation of traditional task-by-task ESW assessments but with adoption of the broader definition now used by OPCS; smaller tasks—those under \$50,000--would be dropped from the sample; and QAG should proceed (with a few other modifications as agreed) with a full sample of randomly selected tasks
 - building on the findings of the initial pilot of three Country AAA assessments carried out late in FY01, additional Country AAA assessments should be conducted in a way that would include a sectoral/thematic slice in a larger country or a multi-country CMU
 - an approach to assessment of process-intensive tasks, for example non-lending technical assistance or other informal products, should be developed and tested
- FY04 would then reflect a full new ESW assessment program based on the conclusions of the task-by-task assessment of the FY02 sample, and the experience gained and lessons of the other pilots listed above

8. This report (Chapters III and IV) provides the findings and synthesis of the FY02 sample of individual ESW tasks. It also reports (Chapter V) the findings and conclusions of the three Country AAA pilot assessment (Mozambique, Peru and Thailand), the lessons of a pilot review of four financial sector assessments (FSAPs) (Bulgaria, Mexico, Senegal and Tunisia) and of a review of 11 'Other ESW Products' (conferences/workshops, country dialogue and consultations), not previously classified as ESW. A review of actions proposed in the Fourth Synthesis Report (Chapter VI) and Conclusions and Recommendations (Chapter VII) follow.

II. OBJECTIVES AND APPROACH

9. As part of the continuing effort to improve the quality of the Bank's operational work, QAG carried out the fifth task-by-task ESW Assessment from July to November 2002. This assessment had three objectives:

- to promote accountability for the quality of ESW;
- to identify systemic strengths and weaknesses in ESW, and promote appropriate changes in policies and procedures to improve the impact and effectiveness of future ESW; and
- to draw lessons from experience, including identifying examples of "best practices" and weaknesses, so that the Bank continues to build on its strengths as a knowledge-based institution.

APPROACH

10. Quality of ESW assessments are limited-scope, real-time reviews of the overall Bank effort in the previous fiscal year. Assessments are conducted by panels composed of experienced senior staff and consultants. They are based on a random sample of ESW tasks. (Tasks and panelists in FY02 are listed in Annexes 1 & 2. The FY02 Approach Paper is included as Annex 3).

11. In FY02, 228 tasks remained as the sampling universe from a total of 420 ESW tasks reported in SAP as 'delivered-to-the-client', after the standard process of eliminating tasks costing less than \$50,000, conferences, consultation, and FSAPs. (Due to the complex and specialized character of FSAPs, QAG agreed with the Financial Sector Board to conduct a pilot review of four FSAPs in FY02. See Chapter V). A random sample of 75 tasks (33% in number and 43% in cost) was then selected for assessment (Box 2.2).

BOX 2.2 FY02 ESW SAMPLE

The sample for the ESW02 exercise was stratified by cost and Region, to cover better those tasks on which the Bank invested more resources, including trust funds, and to ensure Regional representation. Three strata by cost were used: (i) >\$500,000—100% of tasks sampled, (ii) \$250,000-500,000 – 50%, and (iii) \$50,000-250,000 – 25%. No tasks less than \$50,000 were counted in the universe or sampled for FY02. Minor adjustments were made to ensure a minimum Region sample size of eight tasks as well as to avoid TTL overload. This process yielded the following composition within the 75 tasks sampled:

Region	Sample	Universe	Network	Sample	Universe	Cost of Task	Sample	Universe
AFR	9	29	ESSD	7	25	>\$500,000	6	7
EAP	10	33	FSE	1	6	\$250,000-500,000	23	45
ECA	22	65	HDN	17	37	\$50,000-250,000	46	176
LCR	16	53	PREM	38	102			
MNA	8	19	PSI	7	32			
SAR	10	29	OPCS	5	26			
TOTAL	75	228	TOTAL	75	228	TOTAL	75	228

12. **Methodology.** The assessment methodology was essentially that employed in previous years, with the exception that panelists were encouraged to seek feedback, to the extent appropriate and feasible, from counterparts and other external stakeholders. It used two types of assessments—*Rapid* and *In-depth*. All of the tasks went through a *Rapid* assessment, normally using a two person panel (a generalist and a subject specialist). Tasks judged to require more thorough examination were subject to an *In-depth* assessment by a more customized panel. *Rapid* assessments involved structured discussions primarily with the task teams. Others, including managers, could also be consulted. For the *In-depth* assessments there was a more wide-ranging approach. In these cases, the panel *always* held discussions with the managers and other internal reviewers and sought additional feedback from country counterparts, key donors and other stakeholders.

13. Panelists assessed the *Overall Quality* of tasks using a questionnaire structured along four quality dimensions:

- *Scope and Strategic Relevance* looks for clarity of objectives, appropriateness of the scope, relevance to country circumstances and well-defined audience;
- *Internal Quality* seeks well-defined issues, sound analysis, use of empirical evidence, and the logical progression from analysis to conclusions and recommendations;
- *Dialogue and Dissemination* explores communication and participation aspects of ESW work, the task's contribution to policy dialogue, extent of client participation, capacity-building, and partnership with other donors; and
- *Likely Impact* assesses leading indicators of possible impact on the client, on the Bank and on the wider development community.

14. ESW tasks were also assessed for the quality of *Bank Inputs and Processes*, i.e., the quality of support and oversight provided to the task team by line management and the Bank as a whole. These processes were disaggregated into three broad categories:

- Resource Management Issues—elapsed time, task cost, efficiency of resource use, and skills mix of the task team;
- Managerial attention to quality—at entry and during implementation, by management function within the matrix; and
- Other key contributions—Network/Sector Board, peer reviewers, special factors, e.g. country conditions.

15. As part of an on-going effort to refine and improve the methodology, in particular the assessment of likelihood of ESW impact, several adjustments were introduced to the questionnaire in FY02. (See Annex 3, para. 3 for a listing of all changes). The most significant changes were in the introduction of a section on task development objectives (DOs) and task context. Tasks teams were requested to complete a short questionnaire that served to help identify the broad task DOs within nine possible categories⁵. For those DOs assigned a high priority by the team, teams were also asked to identify leading indicators of task success. In addition, teams were asked to provide an indication of the contextual

⁵ The classification of DOs used in the questionnaire was: poverty alleviation; structural and sector policy reform; private sector development; institutional development/capacity building; human development; environmental sustainability; infrastructure development; macroeconomic management; and other (to be specified).

circumstances/factors that can influence the quality of a task⁶. For their part, panels were asked to assess the likelihood of achieving the DOs identified by the task team, for example meeting the success indicators identified by the team for high priority DOs. Panels were also asked to comment on the task context as described by the task team if they viewed the context as meriting particular attention in determining their assessment. Other changes included greater use of external members in panels—11 panelists in FY02 versus four in FY01. And, whenever feasible, panelists consulted clients directly as part of the assessment process.

16. **Ratings.** Panelists exercised judgment in making the individual sub-ratings, in determining the relative weights in reaching aggregate ratings for each of the four assessment dimensions, and in arriving at an *Overall Quality* rating. They took into consideration the type of ESW being assessed, the objectives and scope of the undertaking, the resources made available and the DOs identified by the task team as well as the team's own assessment of the specific country context. Panelists started from a compendium of background materials provided by the TTL or obtained from central Bank files. Panelists then formulated a set of questions to be forwarded to the task team in advance of a meeting with the team. These questions typically served as an informal agenda for the interview, an important aspect of the review process. Panels were encouraged to consider sympathetically supplementary oral inputs from task team members and other explanation of the task that supplemented the documentation assembled in advance. Subsequently, panels generally sought to interview counterparts or other stakeholders, as appropriate. The work of each panel was facilitated by a moderator. A key role for the moderator was to help achieve consistency in interpretation and application of the assessment process to assure comparability across ESW tasks.

17. Assessment was carried out on a four-point scale for both overall ratings and each sub-question.

Highly Satisfactory (1) Best practice in several areas and no major deficiencies

Satisfactory (2) Solid piece of work without major deficiencies

Marginal (3) Major deficiencies in some important areas

Unsatisfactory (4) A broad pattern of deficiencies

Of the 75 ESW tasks sampled in FY02, six were assigned an initial rating of *Marginal*. Within this set, three tasks were subjected to *In-depth* assessments (with three teams not requesting this option). Of these three, one task had its *Overall Quality* rating upgraded.

18. Some important caveats on ESW ratings merit highlighting. Much effort has gone into ensuring the consistency and fairness of assessments. The FY02 sample size of 75 tasks provides statistically valid estimates of quality for Bank-wide ESW (within $\pm 5\%$ margin of error at 95% confidence level by dollar cost of task and at 90% confidence level by number of tasks). However, margins of error go up considerably at disaggregated levels such as Regions and Networks because of the smaller sub-samples and the sampling procedure which selects only 25% of the less costly tasks in the sample (those between \$50,000-\$250,000). Year-to-year changes or absolute levels should be interpreted with caution, and for that reason, this report focuses on sustained trends, ordinal rankings and major divergences/strong

⁶ The list of context factors and aspects included: country political context; country institutional context; existing country knowledge base and information systems; status of Bank and/or other donor relations with the country; management changes in the Bank; TTL changes; internal organizational changes; and other issues.

outliers. For example, a difference of 10 percentage points between two Regional or Network ratings is usually not statistically significant.

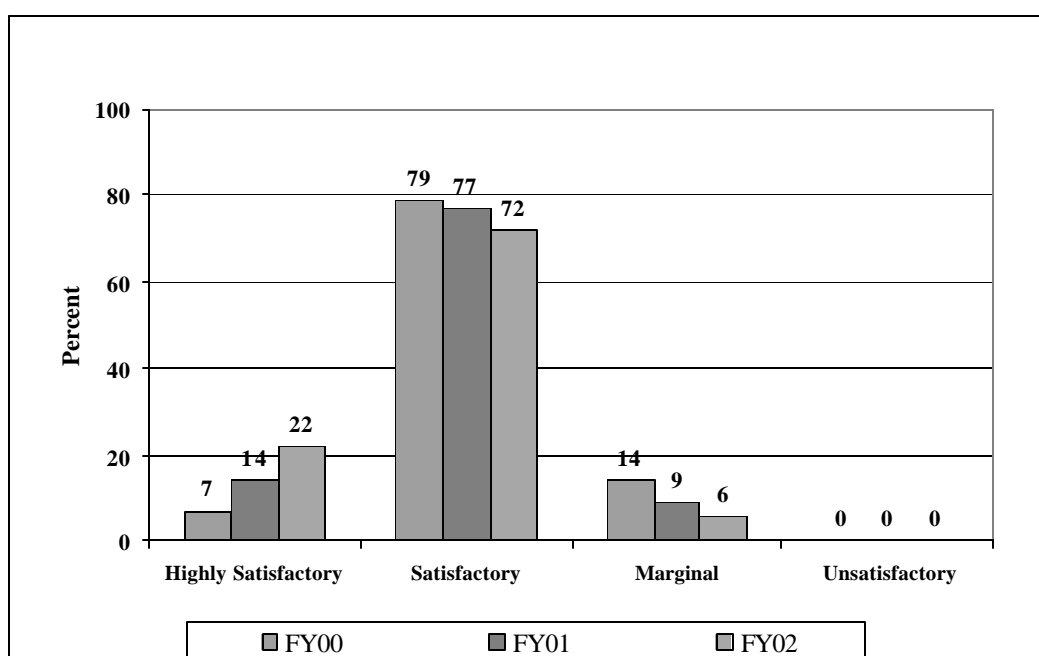
19. Taking into consideration the adjustments made in the assessment exercise covering FY02 -- including the additional check on *Likely Impact* which is afforded by the assessment of DOs as well as the commitment to contact country counterparts -- but also the difficulties inherent in real-time desk reviews, QAG is confident that the overall findings are robust. The consistent approach across now five assessments of quality of ESW also provides a reasonable basis to make valid and reliable inter-temporal comparisons. The entire stock of assessments conducted over these four years (310 evaluated tasks in all) provides a rich data-base for examining quality and performance.

III. ASSESSMENT FINDINGS

A. BANK-WIDE RESULTS

20. The assessment of the quality of ESW in FY02 indicates that the progress achieved in recent years has been consolidated (Figure 3.1). At 94% *Satisfactory or better*, ESW quality remained ahead of the Bank-wide target set at 90%. No task was rated *Unsatisfactory*. Indeed, based on the analysis of the FY02 sample, the gains appear to be more robust than previously identified, as they are associated with a convergence across all dimensions of quality to a quite high standard, accompanied, or perhaps explained, by corresponding improvement in *Bank Inputs and Processes*.

FIGURE 3.1: OVERALL RESULTS



21. Important achievements visible in the FY01 sample were sustained: virtually all ESW tasks continued to be carried out within a well-defined strategic framework; task internal quality was *Satisfactory* in by far the majority of cases; and task teams' efforts to develop dialogue with the client by way of increased participation in the task and effective dissemination of the findings continue to be largely successful. In the FY02 assessment, these achievements are reinforced with, in the view of the panels, a much stronger performance in terms of *Likely Impact*. This reassuring picture of relative convergence across the quality profile is both supported and confirmed by a significant improvement in the FY02 analysis of the standard achieved for *Bank Inputs and Processes*.

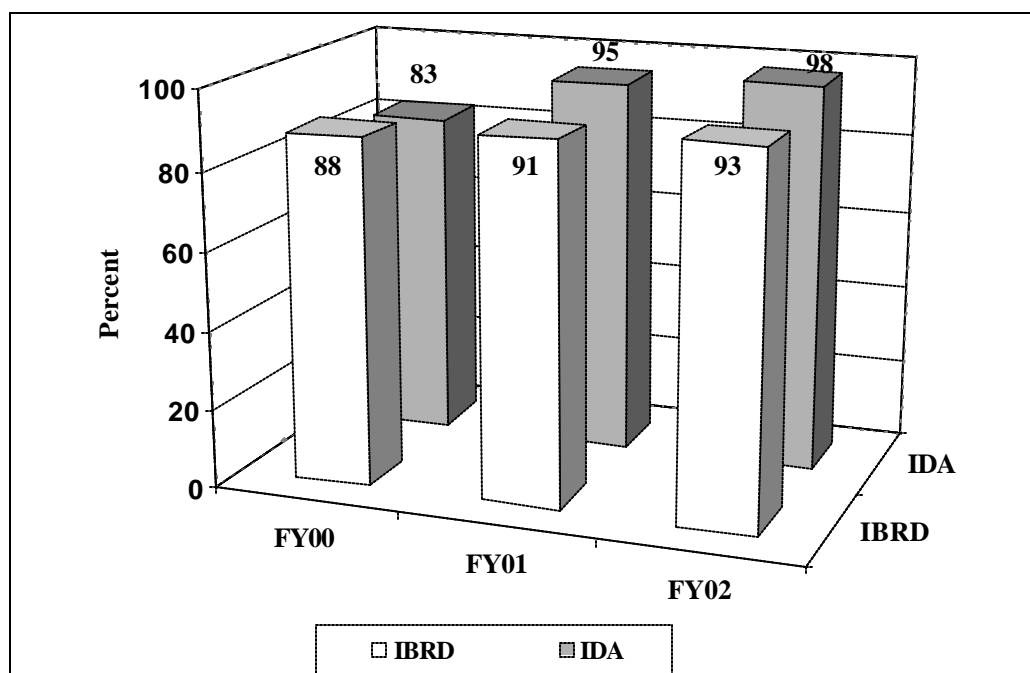
22. Not surprisingly, the share of tasks displaying strong elements of excellence has increased sharply in the FY02 sample. Table 3.1 identifies 16 tasks from the FY02 sample that were rated *Highly Satisfactory*.

TABLE 3.1: A PATTERN OF EXCELLENCE
(Highly Satisfactory on Overall Assessment)

Region	Country	Task Title	Overall Quality	Relevance	Internal Quality	Dialogue & Dissemination	Likely Impact
AFR	Guinea	CPAR	✓	✓	✓	✓	
AFR	Zambia	Public Expenditure Review	✓		✓	✓	
EAP	China	Bank Supervision	✓	✓	✓	✓	
EAP	Vietnam	Development Report	✓	✓	✓		✓
ECA	Belarus	Chernobyl Impact Study	✓	✓		✓	
ECA	ECA	Poverty, Growth and Debt	✓	✓	✓		✓
ECA	ECA	Prague Umbrella	✓	✓	✓	✓	
ECA	Russian Fed.	Labor Market Review	✓	✓	✓	✓	
ECA	Turkey	Corporate Sector Impact of the Crisis	✓	✓		✓	✓
LCR	Colombia	Social Sector Review	✓	✓		✓	✓
LCR	Guatemala	Poverty Assessment	✓	✓	✓	✓	✓
MNA	Yemen	CEM	✓	✓	✓		✓
SAR	India	Improving the Investment Climate	✓	✓	✓	✓	
SAR	India	Maharashtra State Study.	✓	✓	✓	✓	
SAR	Maldives	Public Expenditure Review	✓		✓		✓
SAR	Nepal	CFAA	✓	✓	✓	✓	

23. **IDA/IBRD status.** In the early rounds of assessments (FY98-00), ESW quality was rated as lower in IDA compared to IBRD countries, though the gap narrowed in FY00. Based on the FY01 sample, this picture no longer applied: there was no significant difference identified between the quality of tasks delivered to IDA versus IBRD clients. This situation is sustained in FY02 (Annex 4, Table 6). This reflects improvements in AFR performance and also that in SAR during the last two years. If we adopt a different discriminator—that of low performing countries (defined as CPIA below 3.5, 20 cases in the sample) and other countries (CPIA above 3.5, 55 cases)—as with the case in FY01, any difference detected previously has also disappeared in the FY02 sample (Annex 4, Table 7). As the overall standard of ESW quality has risen, so, it would appear, any difference in quality of ESW service delivered to clients, differentiated either by IDA/IBRD or CPIA status, has essentially disappeared.

FIGURE 3.2: ESW QUALITY BY IBRD/IDA STATUS



* IDA includes blend countries. Regional studies in AFR were treated as IDA. In ECA, regional studies on Eastern Europe were treated as IBRD, and on Central Asia as IDA. All tasks in China and Egypt are excluded, since these major countries shifted from IDA to IBRD during the period FY98-01. The series including them is, however, broadly similar.

24. **Field-managed tasks.** The distinction between headquarters-managed and field-managed tasks has become a rather fuzzy concept as the Bank has decentralized country program leadership to the field while task teams, using modern communications technology, work simultaneously from both locations. Even so, the data-base of the FY02 assessments does furnish some information that allows for a consideration of the possible impact of decentralization on ESW quality. The FY02 sample had 10 tasks that were classified as field-managed (identified as those with the TTL located in the field). All but one of these were rated *Satisfactory*. In confirmation of the FY01 findings, some advantage appears to accrue to field-based task leadership in terms of *Dialogue and Dissemination* and of *Likely Impact* in the FY02 sample (Annex 4, Table 5). Somewhat surprisingly, however, in the FY02 sample panelists have identified a difference with respect to *Bank Inputs and Processes*, with field-based tasks performing less well (63% *Satisfactory* versus 86% for HQ-led tasks). It is not at all clear if this finding, which did not surface in FY01, is an early detection of possible issues with respect to task oversight and quality assurance, as task quality does not appear to have been affected. However, a review of the specific cases which have produced the lower rating for *Bank Inputs and Processes* points to weaker sector management oversight in a few cases, conceivably a more difficult challenge when the sector manager is located at headquarters and the TTL in the field. This question bears watching in future ESW assessments.

25. **Regional Studies.** The FY02 sample contains 10 tasks that are regional in scope, two in AFR, three in ECA, three in LCR, and two in MNA. The balance of assessment of these tasks is also favorable. All of these tasks were judged *Satisfactory* on *Overall Quality*, and comparable in quality to the country-directed tasks across the board. As noted in the FY01 assessment, these studies possess important strengths for the Bank. They can help forge broader partnerships with regional research institutions and donors; they may open up opportunity for assessment of sensitive issues on which countries acting individually might not so easily agree (two cases in the FY02 sample were regional AIDS papers); and, by providing a basis for inter-country comparison, they may foster a competitive spirit that can motivate countries to accelerate reforms.

B. TASK DEVELOPMENT OBJECTIVES

26. As a pilot exercise, in line with the Bank's commitment to a results orientation, the FY02 ESW quality assessment introduced the analysis of ESW development objectives (DOs). There were several reasons for this trial innovation. First, precision in terms of task objectives and their relative importance is a recognized hallmark of good task planning and, at least in terms of a clear statement of objectives, a required feature in a task concept paper. By requesting task teams to classify and rank task DOs, panels were afforded additional information on which to base their assessment. Second, the assessment of *Likely Impact* may be sharpened if DOs are specified in advance, and if teams provide their own view of what would constitute achievement of desired results, the success indicators. Third, a separate judgment by panelists on the likelihood of achievement of tasks DOs, on the basis of available evidence, can be a check on the internal consistency of the overall task assessment. The issue remains, however, that in ESW task planning DOs are not well specified; certainly, neither indicators of success in achieving the desired results, nor interim benchmarks, are typically set down in concept papers. For that reason, prior to the meeting with panelists, task teams were asked to complete a short questionnaire in which they attempted to spell out their task DOs and the corresponding success indicators. The following section summarizes the main results of the panels' consideration of DOs and any lessons that emerge.

27. For a summary of the frequency of respective DOs rated "high priority" by task teams as well as the assessment of the likelihood of achievement of these high priority DOs by panelists, see Table 3.2. (On average, teams identified between two or three DOs for each task as high priority.) The table indicates that structural and sector policy reform was the most frequently cited by task teams as a high priority DO (56% of cases). This should not be surprising since the selection of ESW task is largely driven by the strategic considerations linked to policy reform and institutional development. In that regard, institutional development/capacity building is also quite frequently designated as a high priority DO (44% of cases). So is poverty alleviation (44%).

TABLE 3.2 TASK DEVELOPMENT OBJECTIVES

Development Objective	Rated 'High Priority' by Task Team (%)	Assessed 'Likely' by Panel (%)
Poverty Reduction	44	45
Structural and Sector Policy Reform	56	60
Private Sector Development	13	80
Institutional Development/ Capacity Building	44	61
Human Development	32	71
Environmental Sustainability	4	100
Infrastructure Development	12	56
Macroeconomic Management	20	67

28. With respect to the likelihood of achievement of the respective DOs, panelists considered both structural and sector policy reform and institutional development/ capacity building as likely to be realized in about 60% of cases. As for poverty alleviation, that likelihood of success ratio falls to 45%. This conclusion of higher likely success for policy reform and institutional change, as a result of ESW, may be attributed to panels' finding evidence of potential influence of the task on government decision-making and action. Whereas, in the case of poverty alleviation, indicators for which would move more slowly, the cause and effect relationship between the analytic contribution of an ESW task and the flow of benefits to the poor--which would only be achieved after some considerable lag--is understandably harder to establish. This dilemma does pose the challenge, however, to identify in ESW planning interim benchmarks along a reasonable path to progress in reducing poverty—a so-called causal chain for poverty alleviation. The on-going transition in the Bank to a more results-oriented CAS, with progress reporting based on such interim benchmarks, is a step in the right direction. Part of this agenda would also include stronger commitment to articulate the interim benchmarks and indicators of success for the respective ESW program included in the CAS.

29. In an effort to gain a perspective on the value added of the task DOs and corresponding success indicators, a review was conducted of the quality and utility of the information provided the panels. Several lessons emerge. First, while the request for task teams to identify DOs was useful to panels in providing additional insight into the operational purpose of a task, the quality of indicators of success so identified varied widely--from general assertions ("enhance competition" or "aim to open dialogue") that could not effectively be monitored, to a mere restatement of planned task outputs ("conduct a workshop"). Second, the review of the cases included in the sample also suggests that teams need greater precision in the framing of task objectives and greater commitment to their monitoring during implementation. The FY01 ESW Synthesis Report argued that clear task objectives combined with indicators of success be part of good task planning and incorporated into concept papers⁷. This review of task DOs provides a further basis for encouraging that direction of change. Third, notwithstanding the limitations imposed by the quality of the information provided, the articulation of DOs can be a further refinement in the approach to assessing *Likely Impact*. It provides panelists some additional information on which to base their assessment and in particular to compare the ratings of *Likely Impact*. But an important caution also applies. While the potential utility is clear, this aspect of ESW

⁷ Op. cit. Box 4.1, p. 21.

assessment will only add value to the assessment if the DOs and success indicators are formulated at the stage of task inception rather than retro-fitted for the purposes of task assessment, as is presently the case in the majority of tasks in the sample.

30. Finally, the approach to identifying task DOs as part of a ESW quality assessment process should itself be subject to further review. While reflective of Bank operational objectives, the classification of DOs used in the questionnaire should be revisited in the light of agreed sector objectives and thematic categories. One may even ask if the identification of task DOs should be more open-ended. Importantly, in view of the Bank's corporate agenda, the linkage with MDGs should be more transparent. Certainly, only high priority DOs should be requested of task teams. And an important issue remains whether it is more meaningful and reasonable to identify DOs at the level of thematic or programmatic cluster of tasks within a CAS than at the individual task level.

C. TASK CONTEXT

31. In order to help panelists acquire greater perspective on the task under review, TTLs were asked in the FY02 assessment to identify contextual circumstances that had a significant influence on the task as well as the nature of that influence, whether favorable, unfavorable or of no special significance. About two-thirds of the TTLs indicated that country political context had a significant influence on the task, with more than half of those viewing the influence as unfavorable. Some of these teams identified, for example, a prolonged political uncertainty as an unfavorable influence on the performance of the task. In contrast, in those cases where the teams saw the political context as more favorable, an element of that perspective was often a lively debate within the country on poverty reduction combined with a strong interest on the government's part to carrying out the task.

32. Another aspect of country context that drew commentary from task teams concerned weak institutions, as characterized, say, by lack of effective counterparts, rapid staff turnover or lack of authority of counterparts to participate in and carry out their part of the work. This combination of factors was cited as a constraint on their work by about half of the all TTLs.

33. Existing country knowledge base and the quality of country information systems were also identified as unfavorable factors in the performance of the task in about one third of cases. This judgment by teams was based on their assessment of prior analytical work, undertaken by the Bank and/or other donors, as well as on their judgment of the quality of government statistical and information systems.

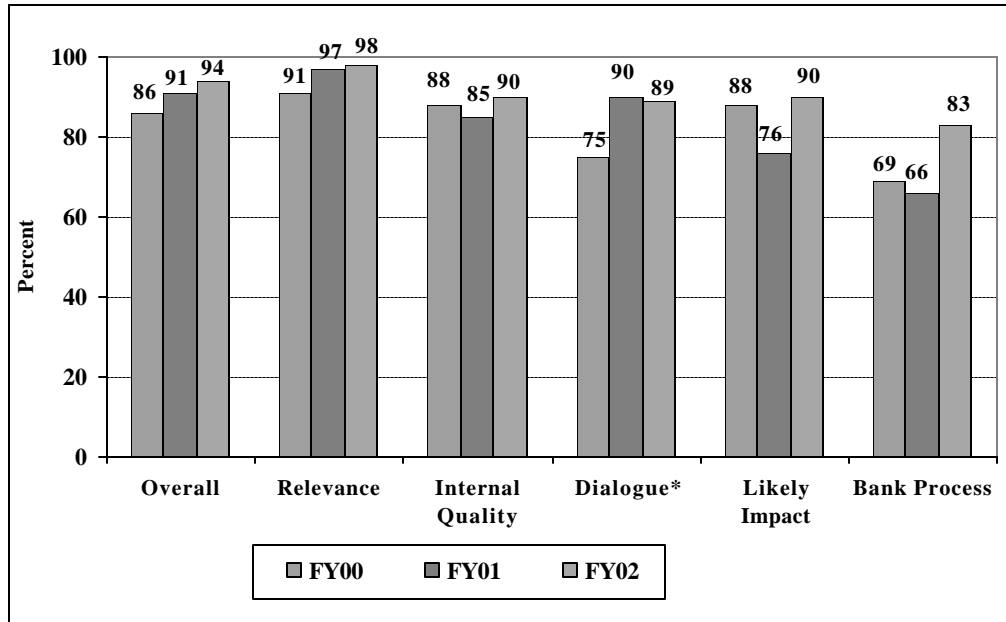
34. As to Bank or other donor relations with the country, more than half the team leaders indicated the quality and status of these relationships as a favorable influence on the implementation of the task. With respect to contextual aspects internal to the Bank, management or TTL changes or internal organizational changes in the Bank during the task implementation seem to have had little or no significant influence on task implementation in the FY02 sample.

D. RESULTS BY QUALITY DIMENSION

35. The assessment methodology is based on the premise that each dimension makes a distinctive contribution to the Quality of ESW. The results for each quality component in FY02 show greater uniformity than in the case of earlier assessments. (Figure 3.3). While some changes are indicative, rather than strictly statistically significant, they also point to a

second conclusion: current ratings for both *Likely Impact* and *Bank Inputs and Processes* are comparable with the standard achieved in the other aspects of quality.

FIGURE 3.3: RESULTS BY QUALITY DIMENSION



* The rating for the dimension Dialogue has a discontinuity from FY00 to FY01.

36. The assessment of quality dimensions, in particular the *Likely Impact*, of tasks in the FY02 sample was reinforced by more consistent effort to make contact with country counterparts and other stakeholders, in cases where this was both feasible and relevant. (For example, contacts with country clients were not attempted in the case of regional studies.) Such consultations were conducted in about 40% of cases, including all cases where panelists viewed direct feedback from the field as necessary to reduce areas of uncertainty in the assessment.

37. These discussions helped clarify client's interest and commitment to a task as well as client's and other stakeholders' degree of satisfaction with the overall process of task implementation, including the outcome (findings and conclusions, and follow-up). Box 3.1 contains a sampling of comments recorded by panelists from client/stakeholder interviews. While there is no common thread or pattern to the feedback obtained through these interviews, both positive and quite critical feedback was obtained. Generally, counterparts had a favorable view of the internal quality of a task. Some critical comments were received about the priority, focus and timeliness of a task; some other commentators expressed concern about less than effective participation in a task, including interaction with and dissemination to non-governmental actors.

BOX 3.1 CLIENT/PARTNER/STAKEHOLDER FEEDBACK TO PANELISTS

Examples of comments obtained through client/partner/other stakeholder interviews:

Government counterpart: “This is a very solid, professional, analytically sound and objective report, which was conducive to a thorough discussion of the issue. For this country, issues of public service were never focused on in the past, and thus this report was a good eye opener...”.

Government counterpart: “The report suffered too many delays, resulting in its coming out at a time when the country has changed fundamentally following the eruption of a crisis...”

Partner institution: “This is the first collaboration with the (partner agency) on this type of task. The experience was so successful that it is being replicated in other countries.”

Partner institution: “The report is indeed very comprehensive, but also too long. The government would have benefited more from a shorter, more focused report. Government officials are still struggling to digest the study. More prioritization and sequencing of recommendations would have been useful. In addition, more ‘how to’ (implement) the findings and recommendations would have proved useful.”

NGO spokesperson: “This is a well grounded and useful report on issues covering relevant parts of the public sector. While some of the findings have been broadly known by the public and by (NGOs) for some time, the sheer association with the Bank is instrumental in giving credence, and thus getting a hearing by the authorities.”

Private sector stakeholder: “To the best of (my) knowledge, the discussion (of the report) has been focused on the government and some technical institutions, and there has been little dissemination and participation of the (concerned private firms themselves) or their association ...”

38. In a number of problematic cases, where panelists had reservations about the rationale for task selection and scope or the degree of local interest, including partner agency participation, interviews provided a firmer basis for the assessment than would otherwise have been possible. Panelists sought of course to balance these comments against information provided from other sources, including donors and the Bank. Notwithstanding the logistical challenge and greater cost associated with obtaining in-country feedback, the process has proved useful and should be continued on a selective basis, particularly for cases where other sources of information (Bank files) are less than conclusive on the questions of client and stakeholder satisfaction and feedback. (See also para. 117 and Annex 6 for panelists’ views of the utility of client and stakeholder interviews.)

39. With respect to assessment of *Likely Impact*, the improvement in that dimension is largely a consequence of greater effort by task teams to create ownership of, and to promote, meaningful changes in government strategy and policies. Panels also noted stronger performance in task teams’ efforts to build support among donors and other partners.

40. The assessment of *Strategic Relevance and Timeliness* based on the FY02 sample confirms the finding of last year that virtually all tasks are now adequately screened at inception with respect to their fit with the client’s interests as expressed to the Bank and with the Bank’s own programming agenda, notably the CAS, and by extension, the MDGs. Panels also note that task teams now provide statements of task objectives, typically included in the

concept paper or equivalent document. However, a new question included in the FY02 review, indicates that those statements are not commonly accompanied by clearly defined indicators of success. (See earlier discussion of DOs; paras. 26-30.) Panelists credited teams with specifying those success indicators in about three out of four cases--which is probably a charitable estimate, as panelists were aware that such discipline in task planning has not typically been requested by managers.

41. The assessment of *Internal Quality* in FY02 also reveals an overall strong performance, with the aggregate result moderately (though not significantly) higher than last. With respect to areas of strength, panels agree that about nine out of ten tasks adequately cover the appropriate areas and issues in relation to task objectives; the critical issues for analysis are well identified in virtually all cases; and, a marked improvement from last year, the analysis is sound in all respects in about nine out of ten cases. Teams are also doing an effective job in terms of tapping into knowledge available within the Bank and elsewhere in the performance of their tasks.

42. An aspect of *Internal Quality* where panelists found further room for progress concerns the framing of conclusions and recommendations. Achievement in this area remains approximately the same as last year: about three out of four tasks are judged *Satisfactory* as to how well teams have managed conclusions and recommendations emerging from their work. Panelists recognize, of course, that not all tasks ought to result in tight prescriptions advanced by the Bank. Teams and their managers may well judge indirect approaches, as well as the use of alternative means (written or oral communications) to transmit messages, as more likely to be effective. Even in recognizing and discounting these cases, panelists continue to critique the messages that are provided clients. In a number of cases, panelists saw that great effort had been devoted, often very effectively, to performing the analysis, but much less attention was paid to synthesizing findings into a reasonable and limited set of priorities for action, including next steps. In not a few cases, panelists wondered if the team considered the elegance of the analysis as their primary purpose, rather than the application of that analysis to strategy and policy. This critique of the lack of balance between analysis and findings applies principally to the version delivered to client, the so-called green cover, which is provided the panel. It is of course true that some of these deficiencies were corrected subsequently—indeed in several cases task teams have communicated to the panels that greater prioritization, selectivity and phasing are part of what they introduced after this question was posed to them in the draft assessment.

43. Under the dimension of *Internal Quality*, the FY01 assessment sought to consider Bank performance on two cross-cutting themes in the ESW program: **poverty reduction** and **gender**. Those aspects were probed again in the FY02 assessment with some modified guidance to panelists, based on the lessons of the previous assessment. The results are as follows. With respect to whether teams are assigning appropriate priority to assessing **poverty aspects** that may fall within their task objectives, the picture is not entirely reassuring. The FY01 assessment reported that consideration of poverty was judged *Satisfactory* in 83% of tasks. In the FY02 assessment, panelists concluded that poverty aspects were adequately addressed in 77% of tasks. This decline may reflect better appreciation by panelists of how to apply this item. But it surely also suggests that there is scope for improvement-- possibly merely in terms of priority assigned by the task team-- in the way that teams exploit ESW tasks to achieve greater insight into poverty issues (Box 3.2).

BOX 3.2 WEAKNESSES IN ATTENTION TO POVERTY REDUCTION IN ESW TASKS

The following are some selected panelist comments on areas for improvement in cases where consideration of/sensitivity to poverty reduction was rated *Less than Satisfactory*:

“Notwithstanding constraints on the information the team had available, it could have been possible to do some distributional analysis of the impact on the poor of various social protection programs...”

“While the concept paper mentions the poverty dimensions of several aspects of the ... economy and of government economic policies, an obvious link being that with high unemployment, the report itself does not delve into the poverty implications.”

“While there is substantial material about the composition of government spending for education, health and social assistance ..., there is no analysis in the report of the likely poverty impact of public spending on these or other sectors of government activity.”

“The report includes very little information on the poverty profile of HIV/AIDS prevalence, and does not systematically discuss the implications of its findings for links between the epidemic and national/local responses and poverty reduction strategies.”

“Rural areas ... have the highest incidence of poverty, but there is variation within the rural districts. While the report discusses poverty issues, its recommendations do not fully reflect the variation”

44. As for attention to **gender**, panelists considered this category appropriate for attention in 76% of cases in FY02 as compared to 61% of cases in FY01 (when a specific question on gender was first introduced into the questionnaire), denoting, perhaps, greater awareness on the part of panels, as a result of more specific guidance on the questionnaire, or, alternatively, a modest shift in the selection of task objectives and areas of interest in the light of the Bank’s Gender Strategy issued in FY01. Whatever the case, the performance of task teams in addressing gender aspects can be strengthened. The results for FY02 at 68% *Satisfactory* are below those of the initial assessment in FY01 of 77% *Satisfactory*. There continue to be weaknesses in both the availability of gender-disaggregated data and quality of gender analysis (Box 3.3). The possible differential impact by gender of actions stemming from task conclusions and recommendations should have been considered more systematically in some cases.

BOX 3.3: WEAKNESSES IN ATTENTION TO GENDER IN ESW TASKS

The following are some selected panelist comments on areas for improvement in cases where consideration of/sensitivity to gender aspects was rated *Less than Satisfactory*:

“The report was gender blind. Sex-disaggregated data were not used for the most part, and important gender aspects such as the effects of migration, labor market participation and performance, and so on, were not analyzed. Likewise, the recommendations did not take account of gender aspects.”

“In the panel’s view the report could have been strengthened through analysis of gender aspects of employment and unemployment, the role of the informal sector ...”

“The report has not adequately addressed gender issues, in particular as they relate to access to and delivery of public services. Data permitting, more could have been done on the gender dimension.”

“Gender aspects are largely overlooked, perhaps because of failure to make effective use of pertinent poverty data broken down by gender.”

45. To provide further insight into both the treatment of gender in ESW and the assessment of gender by panels, a review of the tasks in the FY02 sample and the corresponding assessments was carried out by the PREM Gender Unit⁸. That review identified another six tasks, where in the judgment of the gender team, gender relevant issues could have been incorporated into the analysis, as well as the panels' assessments. Notwithstanding the improvement in guidance to panels, within the set of tasks rated on gender, the gender team also noted further scope for strengthening the consistency of panels in assessing gender aspects.

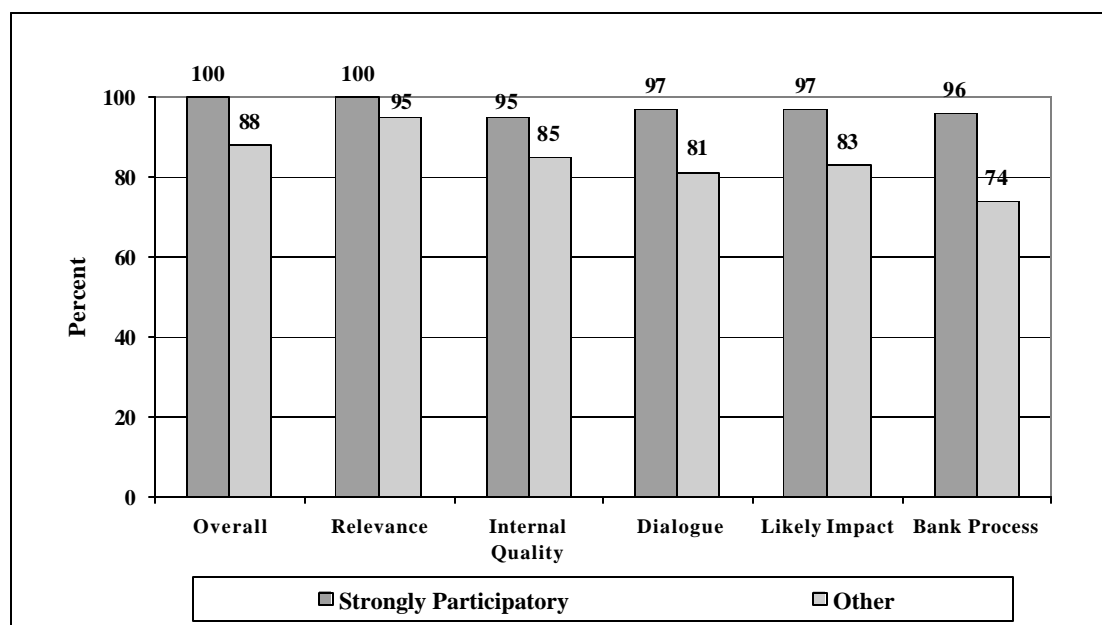
46. In regard to the dimension of *Dialogue and Dissemination*, the FY02 assessment revisited the quality of engagement with clients and counterparts in the performance of ESW. The inclusion of this dimension in FY01 derived from the conviction that the quality of Bank task team's engagement with clients and other stakeholders is indeed an element of ESW quality. That effective local participation in ESW is associated with higher quality work is also substantiated by the results of the FY01 assessment. (Tasks judged as 'highly participatory' were found to be on average of higher quality and more successful in virtually every aspect assessed⁹.) For that reason, the assessment instrument included questions related to how well the client and other stakeholders were involved in defining the task and the issues to be addressed; how active were clients and other stakeholders as participants in the analysis and in the dissemination of the task; and how effective was the Bank team in consulting and sharing perspectives and knowledge, and in engaging in dialogue with clients and other stakeholders, including partners.

47. The results of the assessment of *Dialogue and Dissemination* based on the FY02 sample correspond closely with the findings based on the FY01 sample: 89% *Satisfactory* in FY02 versus 90% *Satisfactory* in FY01. Among areas where further progress was detected are, first, that teams have a clearer strategy of how to engage their counterparts and interact with other stakeholders during the execution of a task; and, second, that task teams are assigning greater attention to activities designed to contribute to client institutional capacity as part of task design and implementation. Both of these attributes suggest a significant commitment on the part of staff and managers to participatory ESW, and thereby promoting client ownership, leading ultimately to higher impact.

48. To probe more deeply into this issue, as in the FY01 assessment, a subset of tasks in the FY02 sample was identified for purposes of analysis as **Strongly Participatory**. The criteria for selection were: evidence of strong client involvement throughout task execution; an iterative process of interaction with interest groups and stakeholders; and, where feasible, use of local institutions. Based on these criteria, 33 (out of 75) tasks in the sample were selected as strongly participatory. These strongly participatory tasks show superior ratings on all dimensions, with a notable advantage in *Likely Impact*, as anticipated, but also, interestingly, in terms of *Internal Quality*. In addition, *Bank Inputs and Processes* were rated much superior for this group. In contrast to FY01 assessment, which identified no important difference in average cost between strongly participatory tasks and other tasks, FY02 tasks designated strongly participatory were on average about 60% more costly than the other tasks (Annex 4, Table 9). Elapsed time was approximately the same.

⁸ *Treatment of Gender in the ESW Quality Assessment, FY02*, PRMGE, December 20, 2002.

⁹ Op. cit., pp. 16-19.

FIGURE 3.4: ESW QUALITY BY PARTICIPATION LEVEL

49. The comparative advantage in performance held by the strongly participatory sub-set adds weight to the conviction that the marginal cost associated with participation is likely to be rewarded in greater pay-off to the task. These benefits arise at task inception as widespread and intensive consultation tends to sharpen task selection and design. They accrue during execution in terms of building client capacity, forging consensus among participants and creating constituencies for change. Another evident benefit concerns the additional insight gained into the tactical approach to managing conclusions and recommendations, including translating findings into policy and programs.

E. RESULTS BY REGION

50. Over the most recent three-year period of these assessments, there has been a convergence among Regions in terms of overall quality (Figure 3.5; Annex 4, Table 1). Although differences are reported across the six Regions, it is important to recognize, both for interpretation as well as follow-up, that below the Bank-wide level the margin of error goes up considerably. Comparisons across Regions (and Networks) must be read cautiously, focusing only on sustained trends or large differences.¹⁰

51. In order to reduce the potential distortion of small sample size and lessen the impact of year-to-year fluctuations, QAG has reviewed the set of assessments for each Region by pooling the results for the most recent two years, FY01 and FY02 (145 tasks in all), capturing thereby the improving trend of that period. The findings reported in Table 3.3 were collected from a careful identification of outliers as found in both the quantitative ratings and qualitative summaries of task assessments. (Features, both strengths and weaknesses, that are common to most Regions are not highlighted).

¹⁰ Many of the more important insights at the level of each Region (or Network) are available from the set of individual task assessments, including write-ups and questionnaires.

FIGURE 3.5: RESULTS BY REGION

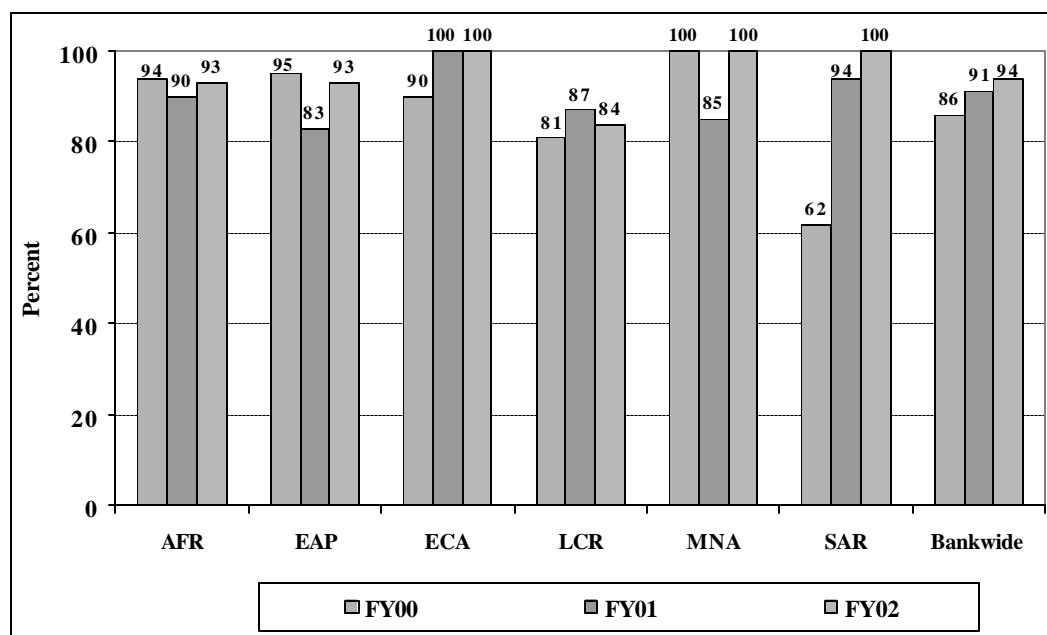


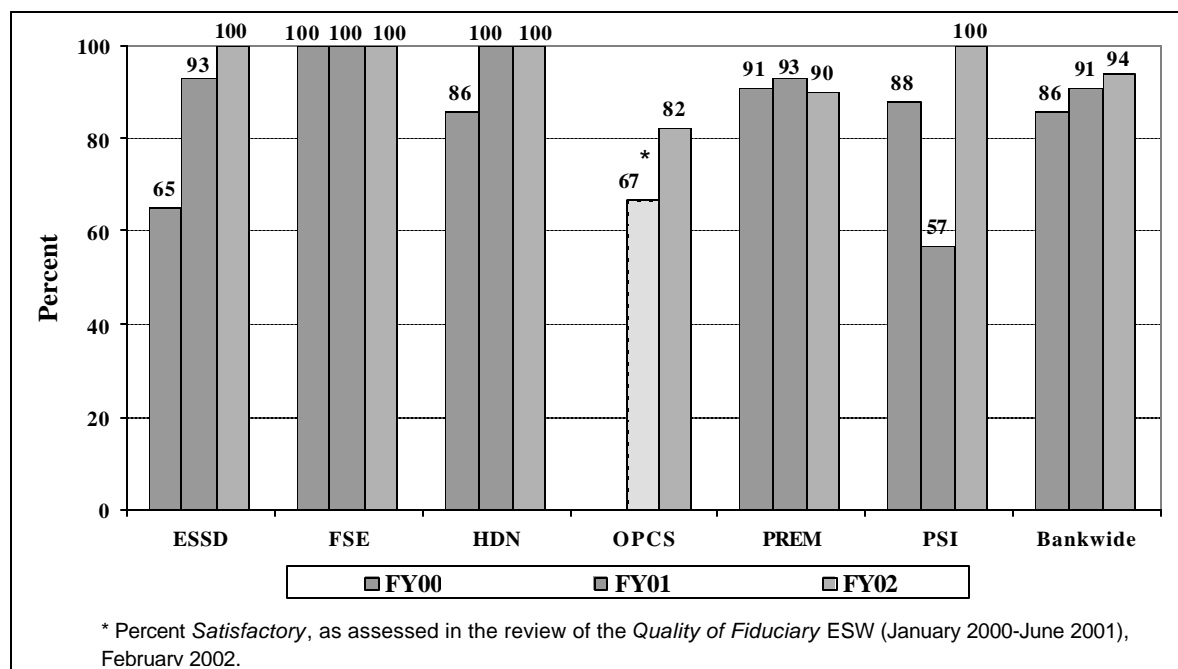
TABLE 3.3. STRENGTHS AND WEAKNESSES BY REGION, FY 01-02

	STRENGTHS	WEAKNESSES
AFR (N=20)	<ul style="list-style-type: none"> Appropriate focus of the task Strategy for engagement and interaction with client and other stakeholder 	<ul style="list-style-type: none"> Use of knowledge from outside the Bank Sector management attention to quality at entry
EAP (N= 22)	<ul style="list-style-type: none"> Client capacity building and knowledge transfer Effective interaction with counterparts 	<ul style="list-style-type: none"> Sensitivity to poverty aspects Management attention to quality at entry and during implementation/dissemination
ECA (N=38)	<ul style="list-style-type: none"> Client participation in defining the scope of the task Management attention at entry 	<ul style="list-style-type: none"> Use of knowledge from outside the Bank Time taken to complete task
LCR (N=27)	<ul style="list-style-type: none"> Strategic relevance and timeliness of task Time taken to complete task 	<ul style="list-style-type: none"> Adequacy of funding Management oversight during implementation/ dissemination Sensitivity to poverty aspects
MNA (N=16)	<ul style="list-style-type: none"> Attention to engaging client and stakeholders Advocacy of change in strategy and policy 	<ul style="list-style-type: none"> Adequacy of funding Use of knowledge from outside the Bank
SAR (N=22)	<ul style="list-style-type: none"> Attention to critical issues in the analysis Quality of empirical evidence Sensitivity to poverty aspects 	<ul style="list-style-type: none"> Client participation in carrying out the analysis and in formulating conclusions and recommendations

F. RESULTS BY NETWORKS AND PRODUCTS

52. As with Regions so by Network affiliation, over the last three years, there has been broad improvement in the quality of ESW¹¹. Network results for FY01 and FY02 have also been combined to increase the sample size for identifying performance and clear patterns. The highlights in Network performance on ESW are presented in Figure 3.6 and in Annex 4, Table 2.

FIGURE 3.6: RESULTS BY NETWORK



53. The same process applied for Regions was followed for the identification of respective Network strengths and weaknesses as shown in Table 3.4. (Features, both strengths and weaknesses, that are common to most Networks are not highlighted).

54. The Bank has now identified a set of *Core Diagnostic* ESW products. These are ESW tasks deemed essential for building a strong analytic foundation for Bank country operations. They focus on four broad areas: poverty; public sector management; fiduciary and institutional capacity; and economic management. The specific products are: Poverty Assessments (PA); Country Economic Memoranda (CEM)/Development Policy Review; Public Expenditure Reviews (PER); Country Financial Accountability Assessments (CFAA) and Country Procurement Assessments (CPAR). The expectation is that the Bank will maintain current (less than five years old) a set of these Core Diagnostic products for each active client.

¹¹ In the FY02 assessment, with a wider eligible universe, including CFAAs and CPARs, the OPCS Network is now included in these assessments. Five tasks were selected in the FY02 sample.

TABLE 3.4 STRENGTHS AND WEAKNESSES BY NETWORK, FY01-02

	STRENGTHS	WEAKNESSES
ESSD (N=24)	<ul style="list-style-type: none"> • Relevance of tasks to Bank's agenda • Adequacy of dissemination arrangements • Adequacy of funding 	<ul style="list-style-type: none"> • Quality of the analysis • Sensitivity to poverty aspects • Consultation with stakeholders during task implementation
FSE (N=7)	<ul style="list-style-type: none"> • Relevance to lending • Use of country knowledge • Recommendations prioritized and selective 	<ul style="list-style-type: none"> • Timeliness of the task in terms of client receptivity • Appropriate scope and coverage of task • Strategy for achieving impact
HDN (N=25)	<ul style="list-style-type: none"> • Client receptivity and timeliness in terms of Bank's agenda • Quality of conclusions and recommendations • Strategy for achieving impact 	<ul style="list-style-type: none"> • Sensitivity to poverty aspects • Adequacy of funding
OPCS (N=5)	<ul style="list-style-type: none"> • Quality of analysis 	<ul style="list-style-type: none"> • Use of country knowledge • Effectiveness of client participation
PREM (N=70)	<ul style="list-style-type: none"> • Quality of analysis • Quality of empirical evidence 	<ul style="list-style-type: none"> • Recommendations prioritized and selective • Management attention to implementation/dissemination
PSI (N=14)	<ul style="list-style-type: none"> • Relevance for Bank's dialogue and lending • Clear identification of critical issues 	<ul style="list-style-type: none"> • Appropriate scope and coverage of task • Arrangements for dissemination • Management attention to quality at entry

55. Overall the FY02 sample includes 30 Core Diagnostic tasks (seven CEM/SSR/DPR, six CFAA/CPARs, eight PAs, and nine PERs). (Annex 4, Table 3). Taken as a group, this set of tasks performed quite well (91% *Satisfactory*) in FY02, though somewhat below the level reached in FY01 (96% *Satisfactory*). In all, over FY01 and FY02, 47 Core Diagnostic tasks have been assessed. This set provides a quite rich data-base for the respective Networks and Sector Boards to consider, both the quantitative ratings and the qualitative comments compiled by panelists¹². In terms of CFAA/CPARs, in particular, FY02 showed marked improvement in overall quality assessment (85% *Satisfactory* in FY02 compared to 67% in the FY01 baseline setting exercise). In view of the frequency of performance of PERs, and their centrality to Bank/country relationship and dialogue, a more systematic review has been conducted of the current crop in the FY02 sample; likewise, a more in-depth review was conducted of the PAs in the FY02 sample—as explained below.

56. **Public Expenditure Reviews.** The FY01 ESW assessment had noted a major improvement in the overall quality of PERs; the percentage of PERs rated *Satisfactory* reached 100% in FY01, compared to 85% previously. This improvement was sustained in FY02, when, again, 100% of the 11 PERs reviewed were rated *Satisfactory*¹³. Maintaining

¹² The set of seven tasks classified as CEM/SSR/DPR performed below the overall average for the overall core diagnostic set. The small size and heterogeneous mix of this set of tasks, in effect a residual category, does not allow for any valid inference or broader generalization with respect to performance. Indeed, the lower rating in FY02 is entirely attributable to two quite different tasks rated as *Marginal*.

¹³ For this special review, 11 tasks in all were selected: nine PERs, as identified by the task teams, as well as two other tasks in the sample which include a sectoral PER and a PER with additional management/ institutional elements. As a group, this set of tasks was quite diversified, both geographically and in terms of emphasis. While most dealt with the size, allocation and efficacy of public spending as a whole, some focused on social expenditures, others paid special attention to public expenditure management; some PERs included an institutional assessment, and at least one was integrated with a fiduciary risk assessment.

that standard can be considered an impressive achievement. Many of the PERs were carried out under conditions which made the task especially difficult. At least a third of the PERs were first such attempts--first tries--in the countries concerned, and lacked the benefits of prior work. In several countries, the political context was unfavorable due to a weak or divided (coalition) government and an absence of consensus on vital issues. Some PERs faced a policy-making and institutional environment still affected by the arduous transition from war to peace or from a socialist system to a market-oriented one, as well as weak local capacity and an even weaker knowledge base. Achieving a *Satisfactory* level of quality under these circumstances reflects well on the competence of the Bank staff and the lessons learned from past PERs.

57. While the progress achieved in the overall quality of the PERs is significant, the detailed sub-ratings as well as the panels' comments in the individual PER assessments based on the FY02 sample suggest that further efforts are still needed to extend the quality and effectiveness of the PERs. The sources of strength identified in the individual assessments include placing issues in a macro--at least fiscal--context and identifying the fiscal pressures well; contributing not just to client ownership but to policy consensus and continuity; assisting in the development of a MTEF; integrating PERs with other work (e.g., CFAA, CPAR); and timeliness, i.e., taking advantage of windows of opportunity for government action.

58. The areas for improvement concern primarily aspects of the task scope, the formulation of findings, conclusions and recommendations, and aspects of quality assurance. First, in the view of panelists, the scope of the PERs was often too broad. This prevented key issues (e.g., institutional weaknesses, the interface between the government and the private sector, the impact of corruption, tradeoffs in expenditures, etc.) from being treated in sufficient depth or at all. The coverage often exceeded the client's absorptive capacity, although sometimes the client was responsible for broadening the coverage excessively. Second, the broad coverage caused, inter alia, the conclusions as well as the recommendations to lack sharpness and cohesion. The recommendations were often too general, almost of boilerplate-type or generic, and rarely prioritized, making up an overwhelming agenda. Seven of the 11 PERs were rated as *Marginally Satisfactory* with respect to whether the recommendations were adequately selective/prioritized. Panels also found room for improvement with respect to dissemination and participation. In a few cases dissemination plans were lacking. And while local participation in the task was very helpful (in terms of ownership and sustainability), when it occurred, the experience with, or incidence of, participation was mixed.

59. And third, panelists commented critically in a number of cases on the inadequate presentation--hurried exposition, poor use of tables and graphs, weak documentation of sources and so on--in the version which was delivered to the client. These negative remarks may point to weaknesses of the supervisory/ advisory role played in those cases on the part of managers or designates, and peer reviewers. With respect to peer reviewers, for example, the record points to an improvement in the case of PERs (as well as overall) in relation to FY01. But there is still some evidence with PERs that they have a greater influence upstream (at the concept paper stage) than downstream, at the decision draft stage, as the draft report is being readied for the client.

60. **Poverty Assessments.** The set of tasks designated as PAs (eight in the FY02 sample) compose a diverse set of tasks, underscoring that poverty ESW has matured somewhat¹⁴. In some countries, a Bank conducted PA is a launch-pad for CDF/PRSP processes. In which case the PA is designed to establish or update a poverty profile and put in perspective the record and achievements of prior or current government poverty policy. In other cases, the PA is second generation, building on a substantial prior investment in poverty analysis. In these cases, the PA seeks to deepen the analysis in selected areas: i.e., the characteristics of pro-poor growth, labor market performance, incidence analysis of public expenditure, or qualitative assessments of social safety net programs. Even though many PAs are now able to build on significant prior work, it is noteworthy that PAs constitute costly exercises. The average cost of the PAs in the FY02 sample was \$330,000, considerably more than the cost of the average task in the sample (Annex 4, Table 9).

61. As in FY01, the set of PAs in the FY02 sample performed well, 100% *Satisfactory* overall and on each of the standard dimensions. This achievement reinforces the conclusion that PAs are selected with keen attention to the strategic rationale: the fit with the cycle of government programming and with the Bank's own country assistance strategy. The quality of analysis, particularly the effective generation and use of empirical evidence, remains a hallmark of PAs. Panels also acknowledged more vigorous effort to develop qualitative information, such as self-perception of poverty and insecurity, as well as user feedback on public services geared to the poor. In addition, in at least two cases, panelists credited teams with an effort to introduce historical context as a source of explanation of the initial conditions—such as land-holding patterns, ethnicity, record of civil conflict and so on—which “explain” the distribution of assets and opportunities, and the political economy of change. As with other tasks, investment in participatory processes appear to be rewarded in terms of possible outcomes and results of the work. Feedback from clients and other stakeholders is typically more positive in cases where they have had a role in design and conduct of the work. At least as important, a strategy for engagement of counterparts has enabled teams to probe into more sensitive aspects—such as geographic and ethnic differences in the incidence of poverty—than would have been possible otherwise.

62. Notwithstanding these achievements, a closer look at the record as revealed by the sub-ratings and extensive panel comments brings to light a more mixed picture. First, with regard to selection of issues for analysis, there is a strong tendency (in both PAs and PERs) to focus task team effort on the social sectors, education, health and other social services, where presumably information is available, including from Bank operations. However, panels commented in several cases that investment in analysis of these sectors may be at the cost of attention to other aspects, notably infrastructure sectors and the rural sector, or on other economic aspects such as labor market participation and performance—for example, lack of attention to real wage trends was cited as a case in point in two PAs. Likewise, migration, an important source of productivity gains for labor as well as a coping mechanism for households, appeared to be relatively neglected in the analysis. Clearly, there are limits to coverage with a fixed budget. But teams could be more thoughtful and more transparent about the choices made.

¹⁴ Six of these PAs established or updated poverty profiles. Among these, the approach to policy analysis differed with some choosing to focus on public expenditure, though institutional aspects were seldom addressed in detail. One task essentially focused on the persistence of inequality. One other task constituted a set of background papers on selected themes.

63. Second, the treatment of public sector management is too often naïve in PAs. There appears to be relatively little attention paid to governance and transparency issues. The leap from analysis of “needs” of selected groups to arguments for switching public expenditure in favor of delivery of services to meet those needs did not for the most part take account of incentives for and capacity of public agencies to respond. When programmatic recommendations are formulated in PAs, they must be based on an informed understanding of the conditions required to ensure effective service delivery.

64. Third, in common with other ESW in the FY02 sample, PAs do not perform well in the movement from analysis to synthesis of findings and conclusions. Indeed, panelists were most critical of this aspect of analysis and questioned if some teams had the maturity and experience, including local knowledge, to do this effectively. Certainly, several PA teams failed, in the view of panelists, to tap effectively into Bank sectoral expertise. Panelists are aware, of course, of the authenticity and legitimacy of local authorities and stakeholders drawing their own conclusions and formulating plans for action. Panelists are also cognizant of the benefit of pure analysis in providing an empirical base for more informed local decision-making. But PAs are also a diagnostic tool for the Bank and in that sense a foundation for dialogue. In some cases, their potential for supporting dialogue appears not to be realized. Several teams had difficulty extracting from the array of findings messages that go beyond the banal. Examples of “conclusions and recommendations” in executive summaries serve to illustrate this issue: “correcting the macro-economic situation remains the first priority”; “there are too many programs of questionable effectiveness”; or “social program priorities for the poor are childcare, health and sewerage”. Even when teams venture more deeply into the programmatic implications of their findings, they failed often to provide a meaningful set of proposals—to identify, for example, relative priorities and enabling actions linked to sensible phasing, as well as clarify likely cost and fiscal implications.

IV. BANK INPUTS AND PROCESSES

65. Panel assessments of *Bank Inputs and Processes* based on the FY02 sample provide a more positive picture than that of previous years. Since the initiation of this series of ESW reviews in FY98/99, panels have assessed the effectiveness of *Bank Inputs and Processes* within a narrow range of 66%-69% *Satisfactory*. On this basis, in about two out of every three cases, *Bank Inputs and Processes*--resources and inputs combined with management oversight provided in support of ESW quality—were found adequate. In contrast, on the strength of the analysis of the FY02 sample, that pattern has begun to break down. The increase in overall ESW quality has been matched, in the view of panelists, by a sharp improvement in the effectiveness of *Bank Inputs and Processes*, assessed in FY02 as 83% *Satisfactory*.

66. Stagnation of ratings for *Bank Inputs and Processes* alongside a gradual increase in the average overall rating for ESW quality had posed an interpretive challenge. If anything, causality would suggest that improvements in ESW quality should be a consequence of strengthening resource inputs and management oversight. The FY01 Synthesis Report examined this issue in detail and the analysis afforded several insights that can largely be summarized as ‘lost opportunities’. Failure to address some longstanding weaknesses in inputs and processes produced ESW output of a quality in a number of cases that was below its actual potential. For example, lack of initial internal Bank consensus (among task, country and sector management); failure to address emerging problems in a timely manner during task implementation; lingering budget issues; lack of early consideration of client participation, provision for dissemination and further attention within the Bank to follow-up on the findings, all of these aspects combined to lower the quality of an ESW task. This conclusion emerged most clearly from a case-by-case consideration of a number of tasks (23% of the FY01 sample) which were rated *Satisfactory* overall, albeit in the lower range, whereas *Bank Inputs and Processes* were judged *Marginal*.

67. What appears to mark the difference in the FY02 sample is the sharp curtailment of the share of tasks that have fallen into the lower end of the *Satisfactory* range. In the FY02 assessment, there are only six tasks (8% of cases, or roughly one third the proportion identified in FY01) with *Satisfactory* on *Overall Quality*, and *Marginal* on *Bank Processes and Inputs*. There were only three cases in the entire sample rated *Satisfactory* overall, with more than one *Marginal* rating on the four dimensions of quality. The picture is consistent with an upward shift in quality of the entire distribution of ESW tasks and, correspondingly, with a sharp reduction of the tail of lower quality, yet *Satisfactory*, tasks. Rather than contributing to the loss of opportunities to enhance quality, as in FY01, *Bank Inputs and Processes* appear in FY02 to have helped more teams seize opportunities to improve quality.

68. With respect to **Managerial attention to Quality Assurance**, panelists identified several important areas of progress—evidence, no doubt, of greater priority assigned by managers to the ESW portfolio--as well as some areas of continued concern. Attention to quality at entry was judged *Satisfactory* in 85% of cases in FY02, a significant gain over that reported in the FY01 sample, 64%. This improvement applied to both sector and country management. Assessment of management attention to quality during implementation/dissemination also showed a gain to about 83% *Satisfactory* in FY02 as opposed to 76% in FY01.

69. The quality of the peer review process--arguably, for most cases, the responsibility of sector management--improved: 76% *Satisfactory* in FY02 versus 63% *Satisfactory* in FY01. However, a note of caution must be sounded. At 83% *Satisfactory* in FY02, the quality of peer reviewer inputs was reasonably adequate; but at 70% *Satisfactory*, the impact of these peer comments on the final product (the draft delivered to client) was less so. This finding confirms what panelists have also commented: whereas written peer review inputs were more commonly available at the concept paper review, they were less often sought, delivered or recorded at the decision draft review. As noted in the FY01 Synthesis, peer reviewers were not always present throughout the life-cycle of the task and there was frequently no accountability for responding to peer advice when the report was reviewed on nearing completion.¹⁵

70. The use of Quality Enhancement Review (QER) is still a rare event in the implementation of ESW. Only four tasks in the FY02 sample were subjected to a QER review, though the cases identified proved creative in accessing specialized expertise beyond the Bank and, for the most part, the QER was salutary, in the view of panelists, in contributing to quality.

71. **Resource Inputs.** The FY02 assessment also afforded a picture with respect to resource availability and use (Annex 4, Table 9). Average elapsed time from inception to delivery was 15.4 months, a modest increase over FY01, 14.5 months. Average total task cost was \$192,000 in FY02 versus \$186,000 in FY01¹⁶.

72. Panelists assessed performance on time and cost aspects of task management in relation to their own judgment of what was needed and doable in relation to the context and complexity of the task. In assessing the FY02 sample, panelists also acknowledged that, in more than a few cases, implementation was adversely affected by the interruption to travel and work commitments caused by the incidents of September 11, 2001. This factor may have also led panelists to be more lenient in terms of judging the 'reasonableness' of the time taken for task delivery, as panelists concluded on this occasion that the length of time from inception to delivery was reasonable in 95% of cases, much more than the level of 78% reported in FY01.

73. The adequacy of funding was assessed positively in 85% of cases in FY02 versus just 67% in FY01, leading to a judgment of 91% *Satisfactory* in the overall efficiency of resource use in FY02, a slight rise over the share reported in FY01. One aspect of task budgeting that drew negative comment from task teams and panels alike concerned the under-budgeting of task dissemination. Initial budgets too seldom included provision for expected or planned dissemination, creating a need in many cases for teams to seek essential financing late in the process of task implementation.

74. Panelists reported that team skill-mix was adequate (88% *Satisfactory*). In those cases where panelists were critical of the skill-mix, their concerns typically related to a few cases where lack of specialized skills was noted, most often lack of access to sectoral expertise in tasks conducted by generalist economists.

¹⁵ Op. cit., p. 23.

¹⁶ Calculation of elapsed time and average cost for both FY01 and FY02 is based on the universe of eligible ESW tasks. In each case, estimates based on the corresponding sample are quite close. See Annex 4, Table 9.

75. **ESW Reporting and Monitoring.** As in the FY01 assessment, panels commonly identified inaccurate and incomplete task records on the SAP. The quality of information on the SAP was judged *Less Than Satisfactory* in about 30% of cases, contrasted with 33% in FY01. The statement of task objectives was often blank or so terse as to be uninformative; milestones of progress were not completed; success indicators seldom listed; and cost data inaccurate or incomplete.

76. This continuing issue with respect to task reporting derives from several factors— inadequate attention to administrative records, and possibly some inconsistency across Regions in task record keeping—, but it also follows from limited incentives for compliance by task teams. Many, probably the majority of teams, report that they do not make use of the system for their own monitoring purposes. For this reason, task teams do not see the SAP record as having substantive value for them in tracking task implementation progress. Evidence of this is found in the frequent need of panels to seek updating and corrections in the reporting, in particular, of task cost data. A most common source of error is the commingling of task costs with other activities or the converse, the reporting of selected task expenditures under other task codes.

77. Panelists also examined the quality of the Activity Completion Summary (ACS). The ACS is a tool for self-evaluation by the task team, including a statement of task objectives and the extent to which these were achieved, the dissemination process and activities, and any lessons learned. The Bank standard calls for an ACS within six months of delivery to client, with a more complete document required for tasks over \$100,000. Approximating the findings on ACS in FY01, the current round found the ACS *Satisfactory* in only about one half of cases. Most significantly, perhaps, only a handful of ACSs contained findings material to the achievement or lack thereof of success indicators, including client feedback. This widespread failing suggests a need either to create meaningful incentives for task teams and managers to comply with ACS requirements or, alternatively, re-consideration of this vehicle for task self-evaluation. In the absence of the use of viable or preferred alternatives, non-use of the ACS, in particular, as a first tier performance monitoring tool, for TTLs, line managers and Bank-wide systems, contributes to the weakness in tracking and reporting ESW task DOs and their corresponding success indicators.

V. PILOT ASSESSMENT FINDINGS

78. The ESW Assessment Advisory Panel Report proposed both an adjustment to the task-by-task review process and a widening of the scope of ESW assessments in accordance with the broader definition of ESW. In support of that agenda, three pilot learning exercises were carried out: a review of Country AAA in three countries; an assessment of four FSAPs, in cooperation with the Financial Sector Board; and an investigation into the quality of 11 tasks denominated as Other ESW (conferences/workshops, country dialogue, consultations). The main findings and lessons of these three exercises are reported below.

A. COUNTRY AAA ASSESSMENT

79. Country AAA assessments were a direct response to a conclusion of the Advisory Panel on Assessing ESW that additional insights into quality may be gained by assessing an assembly of ESW/AAA tasks within some unifying framework. The most important such framework is the country program, the CAS and where appropriate the CDF/PRSP. With ESW being designed to enhance the analytical base for dialogue and operations, there were obvious gains to be sought... or lost... from the synergies both among individual ESW/AAA tasks and with the larger country program and its associated lending.

80. For that reason, a Country AAA pilot assessment was conducted, in collaboration with volunteer country teams, for three countries, Mozambique, Peru and Thailand. The period covered in each case was aligned with that of the CAS in place. Indeed, the countries were selected so that the findings would be available to feed into the preparation process for the new CAS. Each Country AAA assessment was implemented by a panel of two-three persons with extensive Bank experience. The panel reviewed country level and ESW/AAA task documents across the full CAS cycle, supplemented with interviews, both at Bank headquarters and in the countries concerned, with client counterparts and other stakeholders, including donors. The AAA assessments used a country/CAS-wide strategic perspective applied on top of a simplified version of the task level ESW assessment methodology. The departure from the traditional task-by-task approach involved the addition of questions and processes to enable the panels to consider the broad strategic thrust of a set of AAA tasks grouped under one theme or inter-connected set of objectives, as well as the synergies among tasks. The processes also explored the country level allocation of resources for the AAA program, looking at budgets and work programs. Each assessment team prepared a single integrated country report composed of findings and recommendations related to the quality of the country ESW program and its management within the CAS context. Given the pilot context, they separately noted lessons regarding the assessment approach and methodology.

81. **Findings.** With respect to the overall quality of the Country AAA programs, the assessments found all three to be *Satisfactory*. For the purpose of the assessment, the panels found it convenient to group the AAA tasks around the major themes/goals of the CAS. Panels assigned ratings to the themes as well as the individual tasks. Results by theme and by individual AAA-ESW task ranged from *Highly Satisfactory* to *Marginal*. The following discussion highlights the findings in more detail

82. *Focus and Selectivity.* All three panels found a strong link between focus and impact, and significant variation in the sharpness of focus among both country programs and individual activities. One country program especially was criticized for becoming diffuse and fragmented, with a wide range of non-prioritized tasks. All panels advocated greater

selectivity, both in view of resource constraints and in order to maximize impact/results. The panels noted that selectivity did not necessarily mean fewer tasks; it meant, rather, relying on well-chosen tasks that demonstrate client demand, Bank comparative advantage, and prospects for action and continuity.

83. *Synergy and Task Inter-linkage.* Synergy--between the AAA program and the CAS, with and among donors, and among the various tasks--was found to be a key contributor to quality and effectiveness. Panelists stressed the value of synergy not just in substance or content but also in timing. In one instance, the clustering of inherently inter-related activities such as CEM, PER, public sector management review, and CFAA contributed greatly to the collective impact of these tasks. In another case, a concerted focus across several tasks on the condition of the poor and barriers to services provided fresh insight into the mutual reinforcement between government policy and community-based initiatives. Based on the experience reviewed, the pilots pointed out that achieving synergy requires good planning, information, coordination, and good team work and team leadership.

84. *Implementation and Monitoring.* The panels noted that achieving results calls for attention to *implementability* (in terms of Bank capacity, including team skills, and country interest and capacity) at the task planning stage and, subsequently, to *implementation* itself. Rather than advocating rigid adherence to an initial program as set out in the baseline CAS, the panels pointed to the benefits of timely adjustments in response to changes in country circumstances and needs. Indeed, panels concluded that further adjustments in the program were called for in some cases, particularly when country performance changed significantly. To better inform managers of the possible need of program adjustment, panels advocated better monitoring of AAA program implementation in relation to relevant economic/sector developments. To facilitate and institutionalize such monitoring and adjustment, the panels recommended the establishment of monitoring benchmarks, updated periodically, the possible creation of a local (in-country) AAA advisory panel, and the introduction of annual AAA performance reviews, similar to or within CPPRs.

85. *Resource Availability.* Two of the pilot country AAA programs experienced a resource shortage; a third case, arguably, had an excess of funding due to the financing 'windfall' resulting from strong donor/trust fund interest. All three pilot assessments stressed the importance of appropriate funding. In this connection, they pointed out that just as important as the amount of funding were its timely commitment and predictability. Adequate, assured funding was found to be crucial for, *inter alia*, dissemination activities and continuity of effort.

86. *Donor Coordination.* All three pilots emphasized the importance of effective coordination among donors in the conduct of ESW including jointly sponsored work. The experience with donor coordination in the three countries examined has been largely positive, if variable, and improving but, in the panels' view, more could have been done. The objective should be to extend beyond coordination to collaboration. A perennial challenge is to agree on a division of responsibilities among donors and to respect it.

87. *AAA Outside the Formal ESW Stream.* In all three countries, analytical and advisory activities included a significant amount of non-programmed analytic and quasi-analytic work with informal outputs, as well as work related to project preparation. The consensus among the panels was that such work was useful and much appreciated by the clients, and merited continuation. The panels were uniformly concerned, however, over the transparency of, and accountability for, such activities, and the frequent effective bypassing of Bank processes with respect to internal review, evaluation and monitoring. The panels recommended bundling such activities into larger programming units in order to subject it to the normal task initiation and monitoring mechanisms as well as to an *ex post* assessment of effectiveness.

88. *Dissemination.* Based on extensive contacts on the ground, panels found the experience with dissemination to be variable. The apparent persistence of this problem was of concern to the panels because of the close link between effective dissemination and *Likely Impact*. Moreover, in one case, inadequate dissemination was found to have contributed to a perception of secrecy by both the Bank and the government. In some cases, the resistance to broader dissemination has come from the client, although sometimes the Bank was found to have been excessively responsive to client sensitivities and/or not providing sufficient funding for what is often a resource-intensive effort. The panels insisted, *inter alia*, on the preparation of well-articulated, costed and funded dissemination plans for major tasks at the concept stage.

89. *Information Problems.* The panels confronted a variety of information problems which handicapped their assessment work. More fundamentally the absence of this data also implied that country/sector management had serious gaps in their knowledge of the phasing and resourcing of the AAA-ESW program. The problems ranged from identifying the AAA activities, especially those outside the ESW stream, to estimating their cost and locating the outputs. Often, the panels had difficulty determining the total cost of the country AAA/ESW, especially that part relating to Trust Fund financing, and had to develop *ex-post* rough estimates of their own. The panels' recommendations in this area included improving the system for collecting such information and establishing, in collaboration with clients and donors, an easily accessible, web and library-based repository of studies.

90. Finally, several initial lessons and conclusions emerge regarding the approach and implementation of Country AAA assessments, as distinct from the task-by-task approach. First, it is clear that, in selecting the Country AAA program as the unit of analysis, panels are better able to identify significant variation among countries, themes and tasks, and draw attention to areas of relative weakness (timely adjustment, approach to dissemination, resource availability, etc.) that fall within the Bank's control. Second, quality assessment of tasks grouped within themes combined with direct feedback from counterparts and stakeholders added reliability and validity to the process. Third, the findings of two of the pilots have already proved useful in the preparation of the most recent CAS. Fourth, further testing and development of the Country AAA assessment approach is needed, to explore more complex variants (e.g. a sector in a major country) as well as to further refine the assessment instrument and the process of collecting information.

B. FINANCIAL SECTOR ASSESSMENT PROGRAM

91. A pilot learning assessment of FSAPs was conducted in the fall of CY02. Prepared in coordination with the Financial Sector Board and in collaboration with the IMF, this pilot consisted of assessment of four FSAPs (Bulgaria, Mexico, Senegal, and Tunisia) nominated for their diversity in terms of country context and economic/financial characteristics. Each assessment was conducted by a specialized panel, consisting of, among others, two senior Bank staff who led two panels each, and a former senior IMF staff, who served on all four panels. The standard ESW questionnaire was customized with the assistance of the Bank FSAP team to respond to the special features of this product. Panels conducted interviews with the Bank/Fund team leaders and selected staff as well as interacted selectively with managers. Contacts were also arranged with key country counterparts. The findings of this pilot point without exception to a high standard of technical work for this product, but also note issues with respect to selective aspects as summarized below.

92. **Findings.** All FSAPs were rated *Satisfactory or better*. Indeed, on the basis of this initial review, the FSAP work carried out jointly by the World Bank and the IMF is generally of very high quality. This view is also shared by key counterparts in the four countries who expressed a high degree of country ownership, supported by evidence that the authorities are following up this work and applying it to improve the financial sector.

93. In spite of this evidence of FSAP quality, including *Likely Impact*, the pilot identified a need to re-examine some aspects of the FSAP. First, as currently designed, the FSAP has three components: stability analysis, development issues, and the review of standards and codes (ROSCs). The ultimate goal of the program is to promote economic development *with* financial stability, with the ROSCs serving as one of the tools. Yet it is estimated by panels that more than half of the time is spent on the ROSC component, thereby reducing the time for the other aspects, especially development issues. In the view of panelists, the relative priority assigned to the ROSCs within the FSAP should be reconsidered. In planning FSAPs, attention to development aspects of financial systems should be strengthened significantly, particularly in low-income countries. Related to the relative priority among objectives, the assessment also noted that, in order to improve further the cost-effectiveness of FSAPs, there is a need for greater selectivity in the scope and content. This is a tendency that, in any case, has been manifested spontaneously in the more recent FSAP work. As experience has accumulated, FSAP teams have concluded that work on detailed issues of minor significance has taken too much time, has had very little value added, and has detracted from the basic issues, particularly in the development field.

94. Second, the profile of resource use for FSAPs among the four cases reviewed points to high cost of the FSAP program--relative, for example, to Bank ESW--combined with considerable delay in product delivery. In terms of Bank budget, the average cost of the four FSAPs was about \$290,000. Since the Fund absorbs the larger share of the ROSC costs, it is safe to conclude that the average cost of the four FSAPs was in excess of \$600,000, Bank budget and Fund budget combined. At the same time, considerable delay was experienced in delivering the full range of products (FSAP for the client, Financial System Stability Assessment (FSSA) for the Fund's Board, and Financial Sector Assessment (FSA) for the Bank's Board). The average delivery time for the three reports was as follows: FSAP to client (10 months); FSSA to the Fund's Board (nine months) and FSA to Bank's Board (18 months). Delay in delivery of these products is the result of a combination of factors related to the internal review processes in the Bank and the Fund, and the relative priority assigned

to them, notably the low priority assigned by Bank management to the FSA. It is noteworthy that, in the view of the task teams, the Aide Memoire provides the greatest value added to the client because it offers much earlier feedback to clients, typically within two-four months of task initiation.

95. In this light, panelists, as well as task teams, have concluded that this set of outputs should be reconfigured, that a single FSAP ought to satisfy simultaneously the needs of the authorities, the Bank and the Fund. This report could be a finalized version of the Aide Memoire and should be available within a reasonably short time (not more than two or three months) after the second of the two FSAP missions.

96. Third, panels also identified a missing link within the Bank between the FSAP exercise and possible follow-up work on the financial sector. Unlike the Fund, which makes use of this product for Article IV consultations, the Bank has no immediate operational tool for linking the FSAP findings to subsequent Bank operations within the country. To remedy this situation, Bank country management should integrate the FSAP into overall ESW programming through the CAS process and annual updates of the country program. In this way, agreement to undertake an FSAP on a given schedule should be matched by a commitment from the country management team to follow-up the FSAP work in the overall context of country programming. Once an FSAP exercise is completed, Bank country management should seek to ensure that development issues are discussed with the authorities. And, for their part, the authorities in each case should be asked to develop a plan of action. This post-FSAP dialogue may not necessarily lead to technical assistance or lending by the Bank but it should serve to catalyze the efforts of international and national agencies. In addition, such follow-up would have small cost compared with the initial investment, but would help ensure that the FSAP investment will have the highest possible return.

97. Finally, as to the process of quality assessment of FSAPs, the findings confirm that the pilot approach is largely valid, though some further customization of the survey instrument will be made. An important challenge is to exercise very close coordination in the assessment process with the Fund as co-partner in the FSAP program.

C. OTHER ESW ASSESSMENT

98. At the request of OPCS, a third pilot assessment was conducted of so-called “Other ESW” tasks. These activities constitute a diverse group of tasks (diagnostics, update of technical papers, conferences, workshops, country dialogue, and consultations). The purposes of these tasks are also multi-faceted in that they may be undertaken to further country/policy dialogue, share global knowledge, provide technical input into areas where clients lack the technical capacity, or contribute more directly to institutional development. They may be client prepared or Bank prepared. They may also be country specific or span a region. In FY02, the Regions undertook 91 such “Other ESW” tasks at a total cost of \$6.5 million.

99. In parallel to the assessment of FY02 ESW, 12 tasks classified as “Other ESW” and costing more than \$100,000 were selected for a pilot assessment¹⁷. The total cost of 12 tasks that were selected for the pilot review amounted to \$3.5 million, accounting for 54 percent of all “Other ESW” costs in FY02. The average cost of ‘Other ESW’ tasks was \$72,000 Bank-

¹⁷ Initially, the universe of ‘Other ESW’ tasks covered 16 tasks. Three tasks (one after assessment) were later excluded from the sample as they were misclassified or not completed as reported in the SAP. Two other tasks were dropped to avoid TTL overload.

wide with significant variation across the regions. In contrast, the average of cost of tasks that were included in the pilot review was \$294,000.

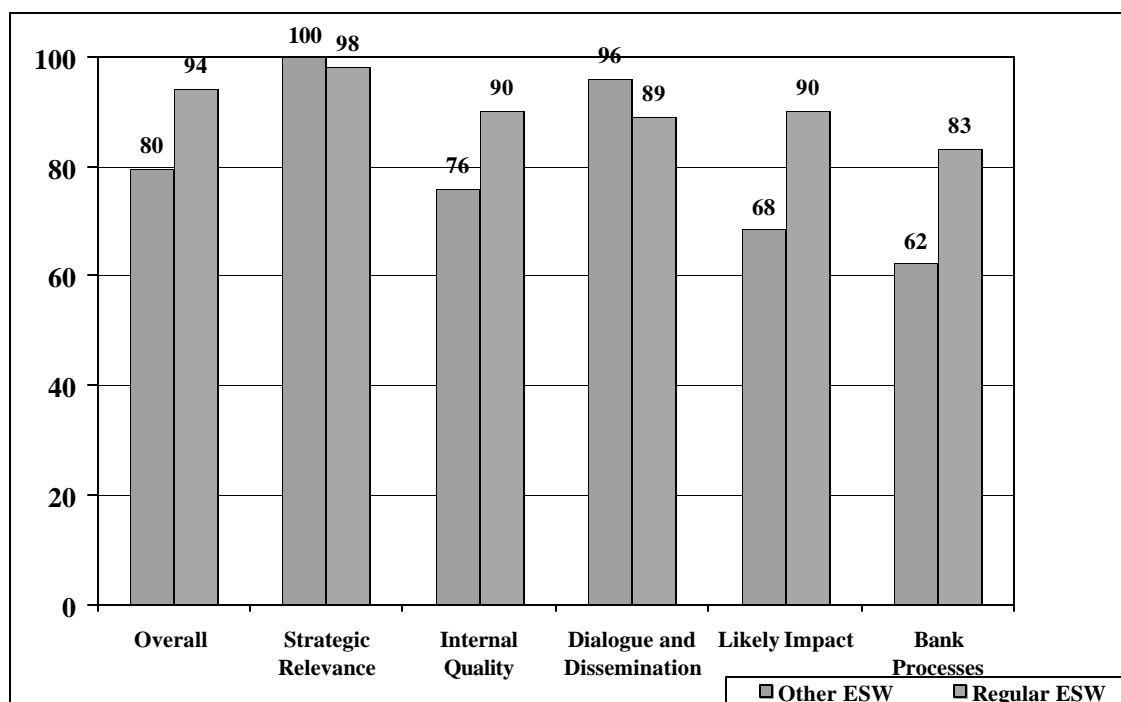
100. All the tasks in this sample have the characteristics that qualify them for an ESW designation: involving analytical content (except one case¹⁸ which was dropped from the sample because of limited analytical content), undertaken with the intent of influencing an external client's policies and programs and owned by a specific Bank unit. The principal difference is that they are often not intended to produce a polished report that goes through various mandated clearances and processes. (In two tasks reviewed, the output was a major report and in the panel's view, they should have been classified *ex-post* as regular ESW. However, the task teams informed the panel that, *ex-ante*, these tasks had envisioned different outputs, which were modified during task implementation.)

101. A panel to conduct a special review of this sample of "Other ESW" tasks was set up. The focus of this initial round of review was on understanding the nature and value of these tasks and developing a methodology for their assessment. The approach to assessment of "Other ESW" was broadly similar to that employed in the regular QAG assessments of ESW. However, the questionnaire used for regular ESW was modified by the panel, as a result of experience gained during the assessments and input received from Task Teams.

102. **Findings.** The overall quality of the 11 "Other ESW" tasks ultimately assessed showed that 80% of the tasks were *Satisfactory* (Figure 5.1). The pilot review concluded that all the "Other ESW" tasks reviewed were rated *Satisfactory* for *Strategic Relevance*, since most of the tasks were integrally linked to country and regional strategy. The strongly participatory nature of most of the "Other ESW" tasks contributed to the strength of these activities in the ratings for *Dialogue and Dissemination*. The results are more mixed for *Likely Impact*, largely due, in the view of the panel, to the uneven attention by sector/country management at the implementation/dissemination stage.

¹⁸ This was a conference which was consciously and appropriately designed as a loosely structured, brainstorming conference, without background papers/analytical notes. That task was considered to be misclassified and was dropped from the "Other ESW" sample after consultation with OPCS and the Region was advised.

FIGURE 5.1: OTHER ESW BY QUALITY DIMENSION



103. On *Internal Quality*, the ratings were lower than found in the regular FY02 ESW sample. The results were similar to those of the regular ESW for the adequate coverage of appropriate areas and identification of critical issues as well as for tapping into the knowledge available in the Bank and elsewhere. However, weaknesses in this dimension were the relatively low scores for selectivity and prioritization of conclusions as well as quality of proposed follow-up actions.

104. The assessment of *Bank Inputs and Processes* revealed substantial weaknesses resulting from the looser internal review processes for “Other ESW”. Management attention at entry and during implementation was considered to be *Marginal* in a number of cases and quite variable in intensity across different tasks. The two tasks rated *Marginal* overall also showed *Less than Satisfactory* management attention at entry and during implementation. Budgetary management of “Other ESW” showed need for improvement. Peer review processes, particularly at the implementation stage, were considered to be *Marginal* in many cases.

105. A key positive feature of “Other ESW” was the frequent use of client prepared documents and in some cases co-task management by the client. In one case the Bank provided analytic advice purely on a demand-driven basis for an initiative started and implemented by a client. In another, the Bank and country agreed on the broad parameters of a multiyear reform agenda, and the Bank provided inputs that were requested for the agreed areas. In yet another case, the Bank conceived the task but had it implemented by a research institution in the country. In a very successful case of coordination, the Bank and client jointly designed and implemented a major conference. In a number of the tasks, the client or local/academic institutes were asked to prepare analytical documents. From the initiating to the completion stages, the task was recognized as the client’s responsibility, with the Bank’s role as one of facilitator and coordinator. Within the sample of tasks selected for review, two

tasks were largely supply-driven by the Bank and thus did not have some of the advantages of effective participation of the key stakeholders. For the most part, the other tasks were conducted with effective participation of stakeholders in the country, and thereby contributed to capacity building and knowledge sharing. These tasks also played an effective role in bringing together the donor community active in the country around a common agenda.

106. The panelists were of the view that the limited sample of 11 tasks in FY02 is sufficiently robust to provide a basis for the above conclusions to be drawn and that the customized guidance questionnaire can be used in future. In addition, the panel identified some aspects of *Bank Inputs and Processes* for these tasks which merit strengthening, as detailed below.

107. *Proposed Review Procedures.* For all larger “Other ESW” tasks, a concept paper or an initiating memorandum should be prepared, with a review process similar to regular ESW. If the Country/Sector or Regional management considers that an advisory group created for the task or an alternative mechanism would be more appropriate than a peer review process, they should have the flexibility to do so, as long as the management decision is recorded. If the task is explicitly part of a multiyear program of thematic, sector, country or regional framework, the concept paper for the umbrella framework and work program would be sufficient. The concept paper should in all cases describe the quality assurance processes and the budget for the task. In the case of tasks with multiple and diverse outputs, the concept paper should also clearly indicate what type of integrative document (e.g. synthesis or summary) will be prepared and what mechanism would be used to ensure consistency of analysis and recommendations across different outputs.

108. The review processes should be flexibly managed to allow variation to accommodate differences in task size, complexity and so on. At a minimum, however, the review processes should be explicitly spelt out at task inception and include reviewers beyond the task team. The choice of review mechanism should not be left to the task manager alone but be approved by the sector/country management. Background papers/analytical papers, major conference design and conference proceedings, for example, should be reviewed.

109. It should be noted that Regions already have detailed guidelines for ESW and AAA, some of which are, or with some modifications could be, applied to “Other ESW”. Most Regions define major ESW tasks as those costing more than \$50,000. The proposed review processes, while broadly similar to regular ESW, have been consciously kept fairly general and flexible to allow for variations for different types of tasks and circumstances, while ensuring acceptable quality standards.

110. A number of tasks in the sample consisted largely or solely of major conferences/workshops. A few additional comments for such conferences are needed. While most of the conferences reviewed (with one notable exception) were well designed and implemented, attention to follow-up activities and their resource requirements needs to be improved. The panel recommends the use of appropriately designed participant feedback questionnaires at the end of the conference. WBI experience in this area could provide a very useful starting point, with necessary adjustments made for the nature of the task. In conferences/workshops, where Ministers and senior officials are the main participants, simplified questionnaires or informal instruments may be more appropriate.

VI. RETROSPECTIVE AND NEXT STEPS

A. RETROSPECTIVE

111. The scorecard on the recommendations of the FY01 Assessment shows quite positive results (Box 6.1). Overall quality of ESW has surpassed threshold standards of 90% *Satisfactory*. In particular, *Likely Impact* of ESW tasks has attained an equally admirable standard in the current assessment. That management has assigned greater priority to achieving excellence in ESW is attested to by the important progress in rating of *Bank Inputs and Processes*. The peer review process shows some tightening but the effectiveness of peer review as a tool for quality enhancement tails off during task implementation. As for the agenda by Region and Network, the overall results are indicative of important progress, though there remains some additional effort. For its part, OPCS has implemented a system of quarterly reporting on ESW; revised country budget norms/benchmarks to strengthen the link between ESW and poverty focus; developed guidelines on the definition and recording of ESW; and worked with Regions and Networks to develop their own ESW processing guidelines. However, further attention is yet required, in particular in the use of the SAP, including the ACS, by the Regions and Networks for monitoring and reporting on ESW tasks. With respect to QAG's commitment to organize a comprehensive review of the ESW assessment approach, the Advisory Panel report has charted the future course of ESW quality assessments, as proposed in the FY01 Synthesis.

BOX 6.1 SCORECARD ON FY01 ESW RECOMMENDATIONS

	Recommendations	Rating*
1	Managers to focus on improving ESW quality and effectiveness	S
2	Chief Economists, SBs and OPCS, clarify guidance and standards for peer reviews	M
3	Regions/Networks should continue to foster a participatory style of ESW	S
4	Regions/Networks should pay attention to the following specific concerns/needs:	S
	<ul style="list-style-type: none"> • AFR/ESSD-- internal quality of ESW tasks • LCR—quality assurance during implementation/dissemination • MNA—attention to poverty and gender aspects and ensure adequate funding • SAR—quality at entry • PSI—ESW standards and processes 	<p>S</p> <p>M</p> <p>S</p> <p>HS</p> <p>S</p>
5	OPCS continue to improve Bank-wide guidance and systems for ESW budgeting, monitoring and reporting	S
6	QAG to conduct comprehensive review of ESW assessment approach ahead of the next cycle.	S

*performance on a standard QAG 4-point scale (see para 17).

B. NEXT STEPS

112. As documented in this Synthesis Report for FY02, the overall quality of ESW has improved substantially, with progress manifest specifically in *Likely Impact*. Both Regions and Networks have attempted to incorporate lessons of good practice in ESW. Importantly, this series of annual reviews has helped focus attention on *Bank Input and Processes* and has contributed to improved management and staff accountability for the use of ESW resources. The maintenance of higher quality standard for ESW, documented in this FY02 review, combined with the initial results of the pilot studies, confirms the wisdom of the general direction for ESW assessments proposed in the Advisory Panel report. Accordingly, QAG proposes now to move ahead in implementing that agenda (para. 7).

113. On the basis of the initial experience in assessment of country ESW, QAG proposes to conduct additional county pilot assessments. Likewise, assessments will be carried out of sectoral/thematic slice of ESW/AAA in some larger countries or grouping of countries within one Country Management Unit. In updating the guidance for ESW/AAA coding and recording in January 2002, OPCS created a new category of Network ESW, though these have not yet been brought into the framework for quality assessment. A targeted sample of Network ESW will be selected for assessment.

C. FEEDBACK ON THE ASSESSMENT PROCESS

114. In order to obtain feedback on the process of assessing the sample of FY02 tasks, a post-assessment survey was carried out with both task team leaders and panelists. 40 TTLs out of 75 (53%) and 40 panelists out of 108 (37%) responded to the surveys. The results are summarized below. A more limited survey was conducted for the FY01.

115. **TTL Survey.** Full results of the TTL survey are provided in Annex 6, Table 1. Survey respondents gave the overall FY02 process a rating of 74% *Satisfactory or better*, comparable to the level of satisfaction in the FY01 (71%). Almost a fifth of TTLs considered that the overall FY02 process was *Highly Satisfactory*; more than half rated it as *Satisfactory*; 14% rated the process as *Marginal*; and 11% assessed the process as *Unsatisfactory*. Comments on the overall process ranged from “open and frank interaction with QAG panelists” and “enriching and rewarding” to “unfair process” and “QAG’s methodology is not robust”. The highest ranking was given to the appropriateness of the tone and atmosphere during the meeting (98%), followed by role of the moderator, panel’s preparation for the meeting and the logistics.

116. Compared to the results for FY01 review there was a significant improvement in the extent to which the panelists were judged to have focused on the relevant issues, a modest decline in TTLs seeing the assessment process as an opportunity for learning, matched by a sharpened perception that the assessment is more an accountability exercise than a learning experience. The TTL survey showed no significant change in the time (16 hours) for document preparation. The average time for post-panel follow-up remained the same at 4 hours.

117. **Panelist Survey.** The complete results of the panelist survey which was conducted for the first time are found in Annex 6, Table 2. While most of the questions were rated more than 90% *Satisfactory*, only about two thirds of the panelists who responded believed that client feedback added value to the assessment. This seems to confirm the anecdotal feedback from panelists that client feedback proved to be particularly useful in the challenging cases.

118. A comparison of the views of TTLs and panelists on the familiarity of the panels with country specific issues and on panel skill-mix reveal some differences of perception. While 84% of the panelists felt that the panel familiarity with country-specific issues was *Satisfactory*; only 62% of TTLs shared the same view. In the case of panel skill-mix, panelists assessed this as 97% *Satisfactory*; TTLs, at 81% *Satisfactory*, were more critical.

VII. CONCLUSIONS AND RECOMMENDATIONS

119. The results presented in this report, in combination with the findings of the preceding ESW assessment syntheses, offer convincing evidence that a quality culture is in the process of taking hold in the performance of ESW. Viewed against the stock of findings accumulated over five years, recent quality gains in ESW appear robust. Notwithstanding this quite impressive achievement, the FY02 task-by-task assessment identified areas for continued attention. In addition, the initial pilots of new approaches, including an enlarged scope for ESW assessments, pointed to important value-added of these ventures. Accordingly the following actions are proposed:

For Regions and Networks:

- Regions should continue to focus on the resilience of quality assurance processes for ESW, including management oversight during task implementation/ dissemination, notably in EAP and LCR, and in the PREM and PSI Network, as well as to provide clear expectations and guidance on the peer review system as it applies, in particular, to the final stages of task implementation, notably the decision draft.
- PREM should make use of the detailed task assessments and summaries of the PERs and PAs in the FY02 sample to provide further feedback to the respective sector boards.
- Regions and Networks should continue to foster participatory approaches to ESW task design and implementation and highlight lessons of good practice to staff.

For OPCS:

- OPCS should encourage Regions and Networks/Sector Boards to improve guidance for incorporating task DOs and success indicators, including interim benchmarks, into ESW task (or program) planning.
- In consultation with the Regions and Networks/Sector Boards, OPCS should arrange to carry out a review of the quality of ESW information in the SAP, and the prevalence of use and quality of the ACS. The purpose would be either to provide workable incentives for use of these tools as designed or to agree on alternative approaches to monitoring and reporting on ESW task progress and results.

For QAG:

- In partnership with Regions and Networks, QAG should move ahead with the next stage of piloting Country AAA assessment approach. An Approach Paper should be presented to Management for mainstreaming in FY04, drawing upon lessons learned from the pilots.

WB158768

M:\ESW FY02\REPORT \Synthesis Report \ESW02 Draft Synthesis Report - march24.doc
March 24, 2003 6:03 PM

ANNEXES

APRIL 1, 2003

ANNEX 1

MAIN CONTRIBUTORS

STAFF DIRECTORS: Praveen Kumar Kadir Yurukoglu

MODERATORS: Robert Drysdale John Sinclair

LIST OF PANELISTS

Alexander Fleming	Fritz Fischer	Maria Emilia Freire
Allen Schick	Gerard Byam	Maria-Luisa Escobar
Amar Bhattacharya	Gershon Feder	Marilou Uy
Amitabha Mukherjee	Giovanna Prennushi	Marlaine Lockheed
Ana-Maria Arriagada	Giuseppe Zampaglione	Maureen Lewis
Anis Dani	Graham Donaldson	Mauricio Carrizosa
Anita Schwarz	Guy Pfefferman	Miguel Schloss
Anthony Churchill	Hans-Martin Boehmer	Mike Stevens
Anthony Kupcis	Helen Sutch	Miria Pigato
Anthony Measham	Helga Muller	Mohsin Alikhan
Arvil Van Adams	Homi Kharas	N. Roberto Zaghera
Arvind Panagariya	Howard Peter Williams	Parvez Hasan
Asad Alam	Ian Hume	Paul Delay
Barbara Kafka	Inder Sud	Peter Bottelier
Basil Kavalsky	Indermit Gill	Peter Miovic
Bension Varon	Ira Lieberman	Ram Chopra
Bernard Hoekman	Jaime Quizon	Richard Ericson
Bryan Falconer	James Douglas	Robert Beschel
Cecilia Valdivieso	James Harrison	Robert Buckley
Cesar Bouillon	James Keith Hinchliffe	Saeed Rana
Christine Allison	Jayant Kalotra	Salvatore Schiavo-Campo
Christine Wallich	Jean-Louis Ginnsz	Samuel Otoo
Daniela Gressani	Jeni Klugman	Santhadevi Meenakshy
David Dollar	Joaquin Gutierrez	Shawki Barghouti
David Rosenblatt	John Holsen	Sri-Ram Aiyer
David Sewell	Jonathan Brown	Steve O'Brien
David Shand	Jose Janeiro	Sue Berryman
David Steeds	Jose L. Guasch	Sue Graves
Donald Winkler	Judith McGuire	Susan Stout
Edgardo Favaro	Kadir Yurukoglu	Tawhid Nawaz
Edward Elmendorf	Kalinidhi Subbarao	Timothy Irwin
Elizabeth King	Kathy Lindert	Uri Dadush
Emmanuel Jimenez	Kin Bing Wu	Valerie Kozel
Eugene McCarthy	Laszlo Lovei	Vinay Bhargava
Eugene Stakhiv	Laura Tuck	Vinaya Swaroop
Florent Agueh	Letitia Obeng	Vincent Galbis
Francis Steier	Lynn Sherburne-Benz	Yaw Ansu
Francis Tuan	Maninder Gill	Zafiri Tzannatos

ANNEX 2

LIST OF SAMPLED ESW TASKS

<u>AFRICA (AFR)</u>		<u>LATIN AMERICA & THE CARRIBBEAN (LCR)</u>	
AFR	Skills Development	Argentina	SME Finances
AFR	Measurement of Trade Reform	Argentina	CFAA
Eritrea	Health Sector Review	Brazil	Ceara Human Dev. And Poverty Reduc.
Eritrea	CEM	Brazil	Bahia Health Decentralization
Ethiopia	Public Expenditure Review	Brazil	CFAA-National Level
Guinea	CPAR	Brazil	Maranhao State Economic Memorandum
Senegal	Country Framework Report	Caribbean	Youth Development
South Africa	Sources of Growth-Phase 2	Caribbean	Education Achievements
Zambia	Public Expenditure Review	Colombia	Social Sector Review
<u>EAST ASIA & PACIFIC (EAP)</u>		Colombia	Poverty Selected Topics
China	Struct of China Agr. Research	Guatemala	Poverty Study
China	BFR-Bank Supervision	Latin America	Regional HIV/AIDS Sector Work
China	Regional Development	Mexico	Technology, Wages And Employment
East Timor	CEM	Mexico	Sugar
Indonesia	Urban Sector Strategy	Paraguay	Poverty Assessment
Lao PDR	PER-CFAA	Peru	Public Expenditure Review
Mongolia	Mongolia PER	<u>MIDDLE EAST & NORTH AFRICA (MNA)</u>	
Philippines	Development Policy Review 2002	MNA	Social Protection Sector Strategy
Philippines	HD Sector Study	Egypt	Poverty Assessment
Vietnam	Vietnam Dev Report/PIP	Egypt	Early Childhood Development
<u>EUROPE & CENTRAL ASIA (ECA)</u>		Jordan	Development Policy Review
Albania	Rural Strategy	MNA	MNA Regional AIDS Strategy Paper
Armenia	Water Resources Plan	Morocco	Public Expenditure Review
Azerbaijan	CPAR	Tunisia	IT Strategy Note
Belarus	Chernobyl Impact Study	Yemen	CEM
Bosnia-Herzegovina	Pub Exp & Inst Rev	<u>SOUTH ASIA (SAR)</u>	
Bosnia-Herzegovina	Local Level Institutional Study	Afghanistan	Country Strategy Brief
ECA	Praque Umbrella	Bangladesh	Poverty Assessment
ECA	Regional Trade Study on SE Europe	India	Local Institutions
ECA	Pov. Reduc Growth & Debt Sustainability	India	Power Reform Strategy Review
Kazakhstan	Gov & Service Delivery Diag	India	Governance Policy Notes
Kazakhstan	CFAA	India	Improving Investment Climate
Kosovo	Poverty in Kosovo	India	Maharashtra Study
Macedonia	Education Sector	Maldives	Public Expenditure Review
Macedonia	Aq. Sector Update	Neopal	CFAA
Moldova	Pub Econ Mat Review	Pakistan	Poverty Assessment
Romania	PER		
Russian Fed	Education Policy Note		
Russian Fed	Children Initiatives		
Russian Fed	Labor Market Review		
Slovak Rep	Social Sector Exp Rev (PER)		
Turkey	Corporate Sector Impact Assessment		
Turkey	Poverty & Crises		

LIST OF NON-SAMPLED ESW TASKS

<u>OTHER ESW TASKS</u>		<u>EUROPE & CENTRAL ASIA (ECA)</u>	
<u>AFRICA (AFR)</u>		ECA	Small Initiatives
AFR	NEPAD	Kyrgyz	Pensions & Soc Safety Note Update
<u>EAST ASIA & PACIFIC (EAP)</u>		Romania	Econ Transition Consul/Conference
Indonesia	Forest PolicyDev & Soc Outreach	Russia Fed	Anti-Corruption Dialogue
Mongolia	Energy Strategy	<u>MIDDLE EAST & NORTH AFRICA (MNA)</u>	
Papua New Guinea	Rural Dev Strategy II	MNA	Public Health Agenda
Thailand	CDP-Governance	<u>SOUTH ASIA (SAR)</u>	
<u>FSAPs</u>		India	Joint GOI/WB Poverty Workshop
Bulgaria, Mexico, Senegal, Tunisia			

ANNEX 3

**ASSESSING THE QUALITY OF FY02 ESW
APPROACH PAPER**

1. While building on experience in assessing ESW accumulated over the past four years, the proposed FY02 Assessment also seeks to implement the recommendations of the recently completed ESW assessment advisory panel (the de Ferranti report).¹ It also incorporates new guidance from OPCS on the definition of ESW and the corresponding coverage of QAG assessments of ESW².

2. The de Ferranti report outlined a phased transition in ESW assessments. As recommended in the report, QAG will implement one further full round of evaluation with a unit of assessment of individual ESW tasks randomly selected (ESW02). In addition, QAG will deepen a pilot program of Country AAA assessments (three were conducted in FY02), and develop and test an assessment approach for sectoral/thematic products. QAG will also develop an approach to assessing “process-intensive” tasks, for example, non-lending TA and PRSP support. Beyond FY03, and subject to lessons emerging from ESW02 and the pilots, the de Ferranti report proposes a transition in future years to a new core unit of account of a Country AAA program in its CAS/PRSP context. This present approach paper covers only one part of this larger agenda for FY03, namely the program of assessment of individual ESW tasks completed in FY02 (ESW02). That program will be implemented over the period mid-June to late-November 2002.

3. Although, as noted, ESW02, will continue with a unit of assessment of individual ESW tasks randomly selected, a number of significant changes are being introduced in accordance with the de Ferranti report. The main changes in coverage and approach are as follows³:

- All tasks under the updated ESW definition will be eligible for inclusion in the ESW02 sample. This would include, therefore, jointly conducted products (such as FSAP), Bank-managed tasks implemented with trust fund financing (such as PPIAF, ESMAP) and Network ESW, all of which were not uniformly included before.
- CFAAs and CPARs completed in FY02, will be integrated into the regular ESW02 universe, as compared to a separate pilot assessment in FY01.
- ESW02 will sample only tasks costing in excess of \$50,000 total cost. The assessment exercise will be conducted no sooner than four months after task delivery to client.

¹ See *Report Of The Advisory Panel On Assessing Economic and Sector Work*, dated May 7th, 2002.

² ESW is now clearly defined as an activity that involves analytic effort, is undertaken with the purpose of influencing an external client's policies, and is owned by a Bank unit. Under this definition, ESW covers five classes of output: country diagnostic reports; country advisory reports; regional reports; other products (policy notes, conferences and other events); and process tasks (country analytical work that does not generate separate products).

³ Other, more minor, modifications are being incorporated into the survey instrument and into the assessment process based on feedback from TTLs and managers.

- Whenever feasible, the assessment process will include direct client input as part of the assessment process in order to strengthen, in particular, the basis for panel assessment of *Likely Impact*.
- Panels will be provided more explicit contextual information on country circumstances, broader country programmatic aspects, or other considerations, which, in the view of the task team, have a special bearing on the delivery and quality of the task under review.
- Panels will include more external members (e.g. other donors, NGOs) in order to bring a stronger partner/stakeholder perspective into the assessment process.
- TTLs will be asked to explicitly define their key objectives of the task, with a focus on impact(s) sought and leading 'success' indicators identified
- In the interest of objective and robust assessment ratings, panels will be encouraged to use the ratings scale in strict accord with the definition of each rating category.
- Finally, QAG will reinforce and expand its activities to promote learning and sharing of findings with staff and intermediate management.

WHY QUALITY ASSESSMENTS OF ESW?

4. ESW is coming to play an increasingly critical role in the Bank's tool-kit. Bank commitment to the MDGs, the on-going PRSP/CDF process, the emergence of the Knowledge Bank, and growing sensitivity to ownership and participation, all point to the critical role of good policy based on sound analysis. ESW provides the foundation for the Bank's policy dialogue with clients, the development of country strategies, and the formulation and implementation of effective lending programs. ESW is also a key instrument for building institutional capacity, especially when it is undertaken with local partners. For all of these reasons, the quality of ESW is important to the Bank, to its clients and to the wider donor community. Recognizing the resources involved, Bank management views assessment of ESW quality as a core responsibility in the interest of accountability and learning. (The Bank is expected to spend \$72m on ESW in FY02, compared to \$106m for lending and \$137m on supervision.)

5. The ESW02 Program of Quality Assessments is designed to:

- *Promote accountability* by providing management and staff with robust quality indicators based on credible, independent assessments.
- *Identify systemic strengths and weaknesses* in ESW and promote appropriate changes in policies and procedures relating to ESW, including its management and review; and
- *Learn from experience* by disseminating findings and conclusions, including examples of best practice.

The broad results and recommendations of ESW assessments are reviewed with Management and CODE via a Synthesis Report. (The last report covering FY01 was issued in March 2002 and is available on the QAG website.)

HOW WILL THE ASSESSMENTS BE DONE?

6. **Coverage:** The QAG Assessment is based on ESW as it is recorded under SAP. Given the new OPCS guidelines and in line with the recommendations of the de Ferranti report, this assessment will reflect a number of adjustments: it will exclude tasks under \$50,000 in total cost but now include Fiduciary ESW, Network ESW, Joint Products and most trust-fund financed ESW.

7. **Quality Dimensions :** Assessments focus on whether we are doing the right ESW, doing it the right way, as well as using the ESW products and processes to achieve optimal impact. They are structured along the following four dimensions:

- *Scope and Strategic Relevance:* Taking into consideration the country context and CAS objectives, this section asks if it was the right ESW to do. Are relationships with MDGs suitably developed? Are the objectives, including the audience, well-defined? Are the expected outcomes and measurable indicators of success identified? Are issues addressed by the task appropriate in view of the objectives? Is the task relevant to the country circumstances and priorities? Is it timely, in terms of country circumstances, receptivity or other key externalities?
- *Internal Quality:* This section looks at the workmanship quality of the task. Are the issues being addressed clearly specified? How good is the analysis and the supporting empirical evidence, both quantitative and qualitative? Does the analysis lead to substantive conclusions? Does it adequately explore the causal chain to results? Are the recommendations operational and consistent with the conclusions? Are they realistic in the light of social, political and administrative constraints?
- *Dialogue and Dissemination.* This section responds to the desire to assess the quality of engagement of the client and other key stakeholders. It explores how effectively they were involved in defining the issues, developing the analysis and managing the dissemination. How effective was the Bank in building ownership with its country-partner? And in capacity-building? This section also assesses the role of other partners, notably donors. It also examines the quality of presentation, in terms of clarity of arguments, adequacy of the summary, transparency/readability for the intended audience.
- *Likely Impact:* This section makes anticipatory judgements about potential impact, in relation to the purpose/objectives and audiences specified by Bank management and the task team at the outset in the concept paper or equivalent. In order to assess *Likely Impact*, the panel uses a variety of qualitative methods that focus upon leading indicators of eventual outcomes, including any self-assessment in the ACS system. Factors for consideration by the panel include: the presence of a sound strategy to achieve impact, engagement of the right target audiences, and appropriate consideration of linkages to policy dialogue and programming. Three possible audiences are considered by the panel, with the balance of priority among them guided by the task team's purpose as stated in the initial task concept paper or equivalent. First, in most cases, the core audience is the country client and target groups within the country; for regional studies, the client may be several countries or regional bodies. Second, the Bank itself (notably its CAS or other key strategic statements) may be a beneficiary. Third, the broader development community may

be a client, or at least an indirect beneficiary⁴. In reaching an assessment for this dimension across these audiences, panelists obtain direct feedback from client representatives and other stakeholders, as appropriate.

8. The assessment also makes a separate set of judgements on *Bank Processes*, seeking to understand the quality of the Bank's own managerial contribution as the ESW was carried out. This component assesses areas such as resource availability (\$, people), cost-effectiveness, the contribution in terms of ideas/guidance from key elements of Bank management (regions, networks, peers, etc). It reviews this contribution at key stages in the task cycle.

9. **Ratings:** The relevance and relative importance of the above dimensions varies with the type of ESW analyzed and the country context. Based on their understanding of the country context and the task-specific goals, panelists exercise judgment in deciding on the relative weights to be given, first to the individual sub-ratings, then among the four assessment dimensions. Such a balanced set of judgments is a central responsibility of panelists.

10. Quality is rated on a four-point scale. Panelists are instructed to use the full range afforded by the scale in accordance with the standards described below. Within each rating category, they are encouraged to use the **comments section** for sub-ratings and the summary **narrative text** to differentiate among stronger and weaker performance.

<i>Highly Satisfactory (1)</i>	Best practice in several areas and no major deficiencies
<i>Satisfactory (2)</i>	Solid piece of ESW and no major deficiencies
<i>Marginal (3)</i>	Major deficiencies in some important areas
<i>Unsatisfactory (4)</i>	A broad pattern of deficiencies

11. For ESW02, the unit of assessment remains the ESW task with its own SAP ID. But the assessment is being conducted in the light of broader programmatic and country considerations. Individual ESW tasks may not be stand-alone pieces; they may be sub-tasks in support of a larger exercise such as a PRSP or else one of a series of small tasks delivered over a period of time. Also, the approach to and depth of analysis may reasonably be expected to differ with country context/circumstances, across different types (sectors, themes) of ESW, and with different task budgets. For these reasons, it is important that panelists, from CAS/PRSP or other sources, understand the larger Bank programming framework and relevant country context. As part of preparation for the assessment process, QAG is asking task teams to respond to a short list of questions on the key objectives of the task, the broader programmatic context, including related Bank activities/reports, i.e. previous ESW, as well as special country circumstances and conditions that may have a significant bearing on the quality and management of the task. Task teams are also being asked to help characterize the nature of client participation in the task. For their part, panelists are requested to summarize their views of the likelihood of achieving task objectives as well as comment on salient country contextual and programmatic factors in the introductory section of the questionnaire. The importance and influence of these aspects in

⁴ This may be especially true for some Network ESW.

relation to ESW quality are also issues that the panel probes with country correspondents. Finally, notwithstanding the effort to conduct a thoroughly technical and professional review, it is important to recognize that the methodology is not designed to provide a definitive judgment on any individual task. For that reason, QAG only reports in the final synthesis broad aggregates and trends, not specific details of individual task shortcomings.

12. **Sample:** The assessment will be based on a stratified random sample of ESW tasks "completed" in FY02. The formal trigger for entry into the sample is the standard SAP milestone, *delivered to the client*. QAG will share the sampling universe with all the Regions and Sector Boards prior to drawing the sample and also the actual sample as soon as it is drawn. Sampling is expected to be in two batches, one in late-May, based on tasks already delivered at least four months earlier, and the second around Labor Day covering the balance of delivery in the fiscal year⁵. Tasks managed by Task Team Leaders (TTL) who have been subject to a QAG evaluation in the previous 12 months will be replaced by other randomly selected tasks.

13. The sample will be stratified by cost and Region. The sampling percentages by cost strata will be, (i) >\$500,000: 100% (ii) \$250–500,000: 50% (iii) \$50-250,000: 25%.⁶ The minimum sample size is being increased to ten tasks from each Region. Based on SAP/BW and CRM data, Regions are promising to deliver about 350-400 ESW tasks during FY02. It is estimated that, after allowing for slippage and ineligible tasks, the QAG sample may include about 80-90 tasks. This is expected to represent about 35-40% of the universe by cost. This sample size should permit a robust assessment of how well ESW dollars are being spent Bank-wide (with less than 5% sampling margin of error at 95% confidence level). (It also permits estimation of ESW performance Bank-wide in terms of number of tasks with a sampling margin of error around 5% at 90% confidence level.) At the level of the larger Regions and Networks, the sampling margin of error should be approximately 10% at the 90% confidence level.

14. **Timing:** The timing of the assessments, usually within 4-8 months after completion of an ESW task, means that in almost all cases it is not possible to judge actual impact. Panels are making an anticipatory judgment based on the workmanship of the report, level and quality of client participation and dissemination, plus direct feedback from client and other stakeholders. A more thorough longer-term assessment of impact would clearly depend upon effective self-evaluations by the Regions/Networks (through ACSs or otherwise) and via OED's Country Assistance Reviews and Sector Impact Evaluations.

15. **Rapid Assessment:** ESW01 involves a two-stage process. In the first 'Rapid Assessment' stage, a panel of typically two members⁷ reviews each sampled task, using the guidance questionnaire as the framework. Their work starts with a review of the task file, a set of background documents assembled by QAG and the TTL. The task documentation

⁵ The commitment to ensure an interval of four months between task completion and task assessment is one of the recommendations of the de Ferranti report. The purpose is to allow greater time for the client to react to the ESW product once "delivered".

⁶ The cost data include funding from all sources recorded in the SAP--Bank budget as well as trust fund and other sources of financing.

⁷ The panel may be strengthened by adding more panelists if the nature of the task demands it (for example cross-sectoral tasks).

covers from concept through to final report, dissemination activities and follow up, including comments by peer reviewers, minutes of decision/review meetings and summaries of discussions with the client. The task team is also requested to provide a short contextual note. Based upon this background material review, the panel prepares a list of issues for discussion during a meeting with the Task Team. Managers may participate in this discussion if they so desire. The panelists also contact appropriate counterparts in the client countries in order to strengthen the panel's judgment on dialogue aspects and *Likely Impact*. (All panel contacts with clients would be arranged in close consultation with the responsible task team and country director/country team.) Following the interview and related discussions, the panelists prepare consensus ratings and a narrative statement. This, and the full questionnaire is shared in draft with the TTL. Where panelists intend to give an overall *Satisfactory* rating to the task but a lower rating to managerial performance, they also consult the managers concerned. The TTL is given five days to comment. The panel is also available to discuss this draft report with the TTL. After the panel finalizes its report, if the final rating is *Satisfactory or better* overall, it is transmitted to the TTL and the line managers by QAG.

16. If the initial overall rating is *less than Satisfactory* (3 or 4), the task normally goes to an In-depth assessment conducted by a more customized panel, unless QAG and the Region concur that a more intensive review would not add value. In addition, the panelists for the rapid assessment may recommend to QAG to convene a panel for an ESW task with special potential for more learning. This approach ensures that no task rating of *less than Satisfactory* is given without adequate consultation with the staff and managers directly responsible, or further panel review.

17. For particularly large, complex or innovative tasks, QAG may opt to go directly to an "Enhanced" mode⁸, with a larger panel and added discussions with Country Director, Sector Manager, etc.

18. **In-depth Assessments:** As with Rapid assessments, the In-depth panel reviews the quality dimensions and Bank processes, but with broader and more intensive approach, notably by adding structured consultations with sector/country directors and peer/QER reviewers. For this stage, panels also obtain wider feedback from clients and other stakeholders within the country to assess their ownership, participation and gauge more completely their views on the *Likely Impact* of the task. The draft report of the panel is discussed with regional staff before being finalized. In addition, the Region (typically via the Chief Economist) is given the opportunity to attach a one-page rejoinder to the final assessment report, focusing on areas of divergent views (and reasons why) as well as on any proposed follow-up.

19. With a view to support staff development and disseminate QAG methodology in the Regions and Networks, they may propose an *observer* (a relatively less experienced staff member) to attend a panel session.

20. **Panel Composition:** The panel, and its members, lie at the heart of the QAG assessment approach. Assessments are carried out by experienced, credible, senior professionals, with relevant technical and/or geographic expertise. Panelists would normally

⁸ For these there would be no second stage review. They would be de facto 'In-depth'.

be active Bank staff level GH or above, retired Bank staff or consultants of comparable experience. To promote diversity in the panels, QAG will this year seek wherever possible to engage a third external panelist from academic or research institutes, from other donors, NGOs, etc., as may be appropriate. QAG also seeks the active support of the Network and Regions in identifying and selecting panelists. To avoid any potential conflict of interest, an individual who has been involved in a particular ESW task, as peer reviewer, manager, or otherwise, or who is currently a staff member of the same Region, may not be selected to serve as a panelist on that task. Panelists are provided guidance and orientation as part of the start-up process prior to panel duty. Equally, QAG provides a **moderator** for each assessment to ensure consistency of approach, including firmness of ratings, among panels, across sectors and Regions.

21. For a *Rapid* assessment, the panel normally consists of three persons composed of a generalist, preferably with knowledge of the country/sector/topic, and two sector/topic specialists as may be required. For *In-depth* panels there are normally three or more members: a country specialist, one of more sector/topic specialists, and a generalist.

WHAT WILL QAG DO WITH THE FINDINGS?

22. **Reports:** QAG issues two types of reports: full assessments, including a narrative section and the full questionnaire for all tasks assessed; and an annual Synthesis Report. Reports for tasks with an overall *Satisfactory* rating are distributed to the TTL, plus the relevant Country Director and Sector Manager. Reports of In-depth assessments are, in addition, sent to the Chief Economist, Regional VP office and the relevant Sector Board Heads.

23. The Synthesis Report provides an overview/analysis of assessment findings, examines systemic issues, highlight patterns of strengths and weaknesses of ESW, and make recommendations for further improvements in quality. The Synthesis may also report the findings and conclusions from any special systemic studies carried out with the benefit of the survey data. The report in draft form is reviewed with participants and other interested observers of the QAG process, then with senior Regional/Network staff involved with quality issues. It is sent to Managing Directors, Regional VPs, Network VPs for review. A final version is discussed with CODE.

24. **Dissemination Channels and Learning:** For the purpose of promoting staff learning through sharing of findings, including best practices, QAG plans to increase activities devoted to dissemination and learning through a variety of channels, including:

- publishing best practice examples on the QAG website.
- providing summaries of lessons and conclusions on the QAG website.
- publicizing best practice examples through interaction with Regions and Networks and Sector Boards.
- working in partnership with WBI and Regions/Networks, participating in targeted staff training workshops, drawing on findings accumulated over the several years of ESW assessments.

WHAT WILL IT COST?

25. Panels for each Rapid assessment of an ESW task are estimated to require an average of 6 staff-days. For In-depth assessments, the requirement will normally be 10 staff-days. Based on a program of some 80 Rapid assessments and 10 In-depth assessments, the total direct assessment time, including moderators, would amount to about 700 staff-days. Additional costs are represented by QAG overhead in management and logistics. The Synthesis document, with its supporting analytical work, is the other core cost center involving senior QAG staff and specialist support. Altogether, the total estimated direct cost for the FY02 ESW program as outlined in this document is roughly \$650,000 (this represents about 0.9% of the ESW budget). As in previous years, QAG will cover the above direct costs of the assessments. Under the new Bank guidelines for cross-support, QAG would reimburse staff for time billed when they work as panelists or on special tasks.

26. These costs do not include the time spent by Regional staff in preparing for and interacting with assessment panels. At an estimated 3 staff-days for each Rapid assessment and 3 incremental staff-days for each In-depth, the direct cost should be around \$250,000. The budget guidelines make explicit provision for these costs in Regional budgets.

27. **WHAT WILL BE THE SCHEDULE OF KEY ACTIVITIES ?**

Issue Approach Paper in final	end-May
Notify Regions/Networks of ESW02 universe	mid-May
Agree with Regions/Networks on ESW02 universe	late-May
Sample 1st batch of tasks, completed by end-February	end-May
First Rapid panels meet	late-June
Sample balance of FY02 tasks*	early-Sept
Panel meetings end	late-Nov
Synthesis report, completed by	end-Dec

* QAG may sample tasks in three batches depending on rate of flow and timing of task completions

ANNEX 4

Table 1: RESULTS BY REGION

(A) BY DOLLAR COST*. % TASKS RATING AS SATISFACTORY OR BETTER

Region	No. of Tasks			Overall Quality			Bank Processes			Relevance			Internal Quality			Dialogue**			Likely Impact		
	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02
AFR	9	11	9	94	90	93	49	70	87	100	100	93	79	80	93	75	92	100	100	64	84
EAP	12	12	10	95	83	93	60	50	57	95	100	100	95	73	71	72	100	93	86	83	93
ECA	16	16	22	90	100	100	84	80	97	90	100	100	90	94	96	65	100	87	100	86	95
LCR	12	11	16	81	87	84	76	77	77	94	100	94	81	87	84	88	87	77	90	82	77
MNA	6	8	8	100	85	100	100	59	80	100	85	100	100	68	100	90	55	100	100	40	100
SAR	10	12	10	62	94	100	39	79	90	72	87	100	89	94	100	74	75	92	54	75	100
Bank	65	70	75	86	91	94	66	70	82	91	97	98	88	85	90	75	90	89	88	76	90

(B) BY NUMBER. % TASKS RATING AS SATISFACTORY OR BETTER

Region	No. of Tasks			Overall Quality			Bank Processes			Relevance			Internal Quality			Dialogue**			Likely Impact		
	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02
AFR	9	11	9	85	85	94	53	67	79	100	100	94	68	70	94	68	95	100	100	66	88
EAP	12	12	10	91	96	88	61	61	62	91	100	100	91	91	69	86	100	88	91	96	88
ECA	16	16	22	90	100	100	83	68	93	90	100	100	90	94	94	66	100	85	100	73	88
LCR	12	11	16	82	81	77	64	68	74	91	100	85	82	81	77	91	81	81	82	82	70
MNA	6	8	8	100	89	100	100	44	93	100	89	100	100	56	100	79	67	100	100	56	100
SAR	10	12	10	78	83	100	65	75	87	87	79	100	91	91	100	81	66	87	69	66	100
Bank	65	70	75	86	90	92	69	66	83	92	95	96	87	84	88	78	87	88	89	74	86

* Total costs including both Bank regular budget and TF money were used for calculating the cost-weighted results, which are cited in the main report.

The number-weighted or results by number are just for comparison with the cost-weighted results.

** The rating for the dimension Dialogue has a discontinuity. In FY01 a reformatted dimension was created, with a broader set of sub-questions rated.

This footnote applies to all tables in this Annex.

ANNEX 4

Table 2: RESULTS BY NETWORK

(A) BY DOLLAR COST. % TASKS RATING AS SATISFACTORY OR BETTER

NETWORK	No. of Tasks			Overall Quality			Bank Processes			Relevance			Internal Quality			Dialogue*			Likely Impact		
	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02
ESSD	13	17	7	65	93	100	61	47	84	78	95	100	76	73	100	60	97	70	89	81	89
FSE	3	6	1	100	100	100	100	91	100	100	100	100	100	100	100	100	77	100	100	68	100
HDN	8	8	17	86	100	100	64	84	97	100	100	100	86	94	93	74	100	100	86	75	94
OPCS	0	0	5	-	-	82	-	-	82	-	-	82	-	-	82	-	-	84	-	-	66
PREM	33	32	38	91	93	90	59	79	72	91	98	97	91	89	86	73	88	90	86	82	88
PSI	8	7	7	88	57	100	86	45	90	95	91	100	88	66	100	88	79	74	91	45	100
Bank	65	70	75	86	91	94	66	70	82	91	97	98	88	85	90	75	90	89	88	76	90

(B) BY NUMBER. % TASKS RATING AS SATISFACTORY OR BETTER

NETWORK	No. of Tasks			Overall Quality			Bank Processes			Relevance			Internal Quality			Dialogue*			Likely Impact		
	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02
ESSD	13	17	7	82	89	100	66	53	84	89	97	100	89	68	100	80	96	65	92	77	82
FSE	3	6	1	100	100	100	100	74	100	100	100	100	100	100	100	100	74	100	100	48	100
HDN	8	8	17	82	100	100	75	78	93	100	100	100	82	88	86	65	100	100	82	80	93
OPCS	0	0	5	-	-	80	-	-	80	-	-	80	-	-	80	-	-	80	-	-	59
PREM	33	32	38	91	92	86	64	70	77	91	98	94	88	85	84	80	88	90	88	79	85
PSI	8	7	7	72	63	100	72	56	79	86	69	100	72	94	100	72	62	73	86	56	100
Bank	65	70	75	86	90	92	69	66	83	92	95	96	87	84	88	78	87	88	89	74	86

ANNEX 4

Table 3: RESULTS BY PRODUCT TYPE

(A) BY DOLLAR COST. % TASKS RATING AS SATISFACTORY OR BETTER

PRODUCT	No. of Tasks			Overall Quality			Bank Processes			Relevance			Internal Quality			Dialogue*			Likely Impact		
	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02
Core Diagnostic	14	17	30	89	96	91	49	87	72	89	100	96	83	100	85	66	82	88	81	79	89
CEM/SSR/DPR	4	6	7	100	86	64	65	71	64	100	100	88	100	100	64	16	59	76	100	59	64
CFA/CPA	0	0	6	-	-	85	-	-	85	-	-	85	-	-	85	-	-	86	-	-	71
PA	5	4	8	86	100	100	64	85	89	86	100	100	86	100	100	77	100	100	100	85	100
PER	5	7	9	85	100	100	27	100	61	85	100	100	71	100	85	86	89	87	55	89	100
Other ESW	51	53	45	85	89	97	71	64	89	91	94	99	89	80	94	77	88	90	90	76	91
Policy Notes	5	8	17	100	86	100	81	74	80	100	86	100	100	87	91	68	86	91	81	86	96
Bank	65	70	75	86	91	94	66	70	82	91	97	98	88	85	90	75	90	89	88	76	90

(B) BY NUMBER. % TASKS RATING AS SATISFACTORY OR BETTER

PRODUCT	No. of Tasks			Overall Quality			Bank Processes			Relevance			Internal Quality			Dialogue*			Likely Impact		
	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02
Core Diagnostic	14	17	30	81	96	88	42	74	79	81	100	93	72	100	83	71	83	88	81	65	84
CEM/SSR/DPR	4	6	7	100	81	74	94	63	74	100	100	92	100	100	74	65	42	82	100	42	74
CFA/CPA	0	0	6	-	-	83	-	-	83	-	-	83	-	-	83	-	-	83	-	-	67
PA	5	4	8	69	100	100	38	48	88	69	100	100	69	100	100	69	100	100	100	48	100
PER	5	7	9	77	100	100	9	100	74	77	100	100	55	100	90	77	90	91	55	90	100
Other ESW	51	53	45	87	89	94	75	63	85	94	96	97	90	78	89	80	93	88	90	75	88
Policy Notes	5	8	17	100	80	100	80	67	82	100	80	100	100	90	86	75	80	86	80	80	93
Bank	65	70	75	86	90	92	69	66	83	92	95	96	87	84	88	78	87	88	89	74	86

ANNEX 4

Table 4: RESULTS BY COST* RANGE

BY NUMBER. WEIGHTED % TASKS RATING AS SATISFACTORY OR BETTER

Cost* Range	No. of Tasks			Overall Quality			Bank Processes			Relevance			Internal Quality			Dialogue*			Likely Impact		
	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02
20-50	7	4	0	100	74	-	69	44	-	100	74	-	100	100	-	100	63	-	100	44	-
50-250	36	39	46	80	92	91	68	68	83	88	99	95	80	81	86	69	89	86	83	75	83
250-500	13	19	23	93	95	96	75	81	79	93	96	96	93	90	92	85	86	92	100	86	93
>500	9	8	6	89	86	100	61	47	86	89	100	100	94	72	100	81	100	100	89	86	100
Bank	65	70	75	86	90	92	69	66	83	92	95	96	87	84	88	78	87	88	89	74	86

* Include costs from both regular Bank budget (BB) and Trust Fund, retrofitted to historic series.

Table 5: RESULTS BY HQ/FIELD MANAGED

(A) BY DOLLAR COST. % TASKS RATING AS SATISFACTORY OR BETTER

	No. of Tasks			Overall Quality			Bank Processes			Relevance			Internal Quality			Dialogue*			Likely Impact		
	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02
Field	17	13	10	84	100	95	56	81	63	92	95	95	93	89	90	63	95	100	76	95	95
HQ	48	57	65	86	89	94	69	68	85	90	98	98	87	84	90	79	89	88	92	72	89
Bank	65	70	75	86	91	94	66	70	82	91	97	98	88	85	90	75	90	89	88	76	90

(B) BY NUMBER. % TASKS RATING AS SATISFACTORY OR BETTER

	No. of Tasks			Overall Quality			Bank Processes			Relevance			Internal Quality			Dialogue*			Likely Impact		
	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02
Field	17	13	10	88	100	88	72	80	63	94	94	88	94	87	76	78	93	100	76	93	88
HQ	48	57	65	86	88	93	67	64	86	91	95	97	84	83	90	79	85	86	94	70	86
Bank	65	70	75	86	90	92	69	66	83	92	95	96	87	84	88	78	87	88	89	74	86

ANNEX 4

Table 6: RESULTS BY IDA STATUS

(A) BY DOLLAR COST. % TASKS RATING AS SATISFACTORY OR BETTER

Funding Source	No. of Tasks			Overall Quality			Bank Processes			Relevance			Internal Quality			Dialogue*			Likely Impact		
	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02
IBRD	32	25	39	88	91	93	76	77	85	94	97	97	86	93	92	80	92	89	92	85	89
IDA	30	41	31	83	95	98	56	75	82	88	97	98	89	86	89	71	88	91	84	75	93
Bank	62	66	70	85	94	95	66	76	84	90	97	97	87	89	91	75	90	90	88	79	91

(B) BY NUMBER. % TASKS RATING AS SATISFACTORY OR BETTER

Funding Source	No. of Tasks			Overall Quality			Bank Processes			Relevance			Internal Quality			Dialogue*			Likely Impact		
	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02
IBRD	32	25	39	91	89	90	77	65	84	95	97	94	91	86	87	86	89	89	93	79	84
IDA	30	41	31	80	91	98	58	68	84	88	94	98	82	84	91	73	87	89	84	72	92
Bank	62	66	70	86	90	94	67	67	84	91	95	95	86	84	89	80	87	89	88	74	87

Note: IDA includes blend countries. Regional studies in AFR were treated as IDA; In ECA, regional studies on Eastern Europe treated as IBRD, and on Central Asia were treated as IDA. All tasks in China and Egypt are excluded, since these major countries shifted from IDA to IBRD during the period FY98-01. The series including them, is however, broadly similar.

ANNEX 4

TABLE 7: RESULTS BY CPIA* CATEGORY

(A) DOLLAR COST. % TASKS RATING AS SATISFACTORY OR BETTER

CPIA Category	No. of Tasks			Overall Quality			Bank Processes			Relevance			Internal Quality			Dialogue*			Likely Impact		
	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02
Low Performing	16	26	20	74	96	94	70	79	86	79	100	94	68	86	86	66	94	96	94	74	94
Other	49	44	55	88	88	95	66	64	80	93	95	99	93	83	92	77	87	87	87	77	88
Bank	65	70	75	86	91	94	66	70	82	91	97	98	88	85	90	75	90	89	88	76	90

(B) BY NUMBER. % TASKS RATING AS SATISFACTORY OR BETTER

CPIA Category	No. of Tasks			Overall Quality			Bank Processes			Relevance			Internal Quality			Dialogue*			Likely Impact		
	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02	FY00	FY01	FY02
Low Performing	16	26	20	77	94	89	63	69	85	84	100	89	70	82	85	71	95	96	93	69	89
Other	49	44	55	90	88	93	70	65	82	94	92	98	92	86	89	81	81	85	87	76	85
Bank	65	70	75	86	90	92	69	66	83	92	95	96	87	84	88	78	87	88	89	74	86

* CPIA categories were based on the overall scores of CPIA. Countries whose overall CPIA scores rated less than 3.5 were coded as Low Performing countries.

ANNEX 4

TABLE 8: FY01-02 RESULTS BY PARTICIPATION LEVEL

(A) DOLLAR COST. % TASKS RATING AS SATISFACTORY OR BETTER

Participation Level	No. of Tasks		Overall Quality		Bank Processes		Relevance		Internal Quality		Dialogue*		Likely Impact	
	FY01	FY02	FY01	FY02	FY01	FY02	FY01	FY02	FY01	FY02	FY01	FY02	FY01	FY02
Strongly Participatory*	27	33	100	100	89	92	100	100	86	95	100	97	100	97
Other	43	42	85	88	57	71	95	95	83	85	83	81	60	83
Bank	70	75	91	94	70	82	97	98	85	90	90	89	76	90

(B) BY NUMBER. % TASKS RATING AS SATISFACTORY OR BETTER

Participation Level	No. of Tasks		Overall Quality		Bank Processes		Relevance		Internal Quality		Dialogue*		Likely Impact	
	FY01	FY02	FY01	FY02	FY01	FY02	FY01	FY02	FY01	FY02	FY01	FY02	FY01	FY02
Strongly Participatory*	27	33	100	100	88	96	100	100	86	98	100	96	100	96
Other	43	42	84	87	53	74	92	93	83	81	78	83	57	80
Bank	70	75	90	92	66	83	95	96	84	88	87	88	74	86

* The criteria for strongly participatory ESW tasks are client-driven, iterative process, use of local institutions, and emphasis on dialogue.

ANNEX 4

TABLE 9: AVERAGE ELAPSED TIME AND COSTS BY GROUP

	No. of Tasks		Elapsed Time (month) ^{a/}		Process Costs (\$'000) ^{b/}	
	FY01	FY02	FY01	FY02	FY01	FY02
ESW Universe ^{c/}						
Region						
AFR	30	29	15.6	15.8	178	159
EAP	35	33	12.2	15.3	283	202
ECA	62	65	14.3	16.4	166	202
LCR	32	53	17.4	14.2	206	181
MNA	20	20	10.1	14.2	129	234
SAR	40	29	15.6	15.7	150	186
Network						
ESSD	42	25	16.3	16.0	257	152
FSE	14	6	17.0	6.3	125	200
HDN	33	37	17.1	15.9	168	191
ISN	0	1	-	15.4	-	320
OPCS	6	26	12.1	12.3	176	91
PREM	87	102	12.7	15.8	191	226
PSI	37	32	13.6	16.9	130	194
BANK	219	229	14.5	15.4	186	192
ESW Sample			Weighted Average			
Total Sample	70	75	15.1	15.0	181	191
Product Type						
Core Diagnostic	17	30	13.4	14.6	244	216
Other ESW	53	45	15.6	15.2	164	176
Source of Fund ^{d/}						
IDA	41	31	14.0	15.1	157	213
IBRD	25	39	17.3	15.1	192	173
CPIA						
Low performing	26	20	13.8	15.5	181	247
High performing	44	55	16.0	14.8	181	174
Participation level						
Strongly participatory	27	33	15.0	14.9	186	249
Other	43	42	15.2	15.0	177	153
Location of Task						
Field	13	10	18.2	12.3	232	171
HQ	57	65	14.6	15.4	172	195

^{a/} Average elapsed time in months between AIS signoff date and delivery to client date.

^{b/} Average costs include both BB and Trust Fund from task initiation to delivery.

^{c/} ESW universe is defined as tasks eligible for FY01 and FY02 ESW assessments .

^{d/} China and Egypt were excluded.

ANNEX 5

QUALITY OF ECONOMIC AND SECTOR WORK
SUMMARY ASSESSMENT RESULTS *

	FY00	FY01	FY02
A. OVERALL QUALITY	86	90	92
1. Scope & Strategic Relevance	92	95	96
2. Internal Quality	87	84	88
3. Dialogue and Dissemination	78	87	88
4. Likely Impact	89	73	86
B. BANK PROCESSES	69	66	83

* All data in this table is number-weighted.

The overall assessment of the quality of ESW is not meant to be a simple average of the assessment for the four aspects of quality listed (which, in turn, are not simple averages of respective underlying factors). The panel should use its judgment in weighing the relative importance of each component and factor in any particular case.

QAG ASSESSMENT GUIDANCE QUESTIONNAIRE

Assessment Rating
 1= Highly satisfactory
 2= Satisfactory
 3= Marginal
 4= Unsatisfactory
 NA= Not applicable

A. OVERALL QUALITY

FY00	FY01	FY02
------	------	------

1. STRATEGIC RELEVANCE AND TIMELINESS

92	95	96
----	----	----

General Guidance: In assessing this section, the panelists should look at the Concept Paper or its equivalent, as well as the larger context as reviewed in the CAS, etc. In some cases objectives are consciously changed midway to reflect new realities. The task should then be assessed with reference to the objectives as they were formally redefined.

1.1 Task objectives

NA	NA	92
----	----	----

a) Clarity of Task Objectives

83	84	92
----	----	----

b) Extent to which the task has clearly defined indicators of success

NA	NA	75
----	----	----

Comments:

1.2 Timeliness of the task in terms of

92	93	98
----	----	----

a) Client receptivity

NA	NA	96
----	----	----

b) Bank's programming agenda

NA	NA	98
----	----	----

Guidance: Client receptivity is gauged in relation to economic or political situation. A decision to undertake the task may be partly based on a window of opportunity within the country - actual or perceived. For some tasks, timeliness may also be weighed in relation to the sequence of Bank or donor related programmatic decisions.

1.3 Relevance of the task objectives to the Bank's assistance strategy for the country

98	99	96
----	----	----

(a) towards initiating/moving the policy dialogue forward

NA	96	97
----	----	----

(b) as an underpinning to strategy development (next CAS/PRSP)

70	90	94
----	----	----

(c) as an underpinning to lending operations

NA	91	96
----	----	----

(d) as an instrument for capacity-building, knowledge-sharing

NA	96	92
----	----	----

Guidance: Not all listed elements may be relevant to each task; score NA where appropriate. The most important reference point is, of course, the CAS under implementation at the time the task was initiated. The panelists should look for a specific justification under CAS objectives - which themselves may be defined very broadly. Apart from CAS, panelists should also be sensitive to explicit guidance of the Country Director..

1.4 Consistency of the task with the Bank's sector/thematic strategy and corporate objectives linked to the MDGs

100	100	99
-----	-----	----

Guidance: The idea is not to penalize innovation or exploration of new directions. But does the case for undertaking this task also square with the broader corporate vision and poverty alleviation agenda of the Bank?

QAG ASSESSMENT GUIDANCE QUESTIONNAIRE

Assessment Rating

1= Highly satisfactory
 2= Satisfactory
 3= Marginal
 4= Unsatisfactory
 NA= Not applicable

	FY00	FY01	FY02
2. INTERNAL QUALITY	86	84	88
<i>General Guidance: While this section flows most "naturally" towards assessment of written products, the spirit behind the questions should also be applied to the internal quality and analytical content within dialogue/processes.</i>			
2.1 Appropriateness of coverage/scope of the task, in view of the objectives (focus, range of issues addressed, areas covered)	NA	NA	87
<i>Comments:</i>			
2.2 Clear identification of critical issues	75	91	85
<i>Comments:</i>			
2.3 Quality of the analysis	79	73	94
<i>Comments:</i>			
<i>Guidance: Look for analytical/methodological rigor expected of similar Bank products. (If specific "product" guidelines exist, were they respected?). Check if the weight given to analysis of individual topics is consistent with their relative importance to the objectives of the task. Give extra credit for state of the art, innovative work. Keep the objectives and size (cost) of the task, specific country context and previously available knowledge in mind.</i>			
2.4 Quality of the empirical evidence	83	86	97
(a) quantitative and qualitative accumulated under the task per se	83	86	96
(b) use of knowledge available inside the Bank - from Network, e.g. anchor thematic groups, DEC, from other Regions, etc.	NA	NA	89
(c) use of knowledge from outside the Bank, notably within the country, (e.g., think-tanks, universities, NGOs, donors)	89	100	93
<i>Comments:</i>			
<i>Guidance: The extent to which the team sought and assimilated information from various sources. Give due allowance for availability of information in the country, and check if the team made sufficient efforts to get to the right information. Keep the size of the task in mind.</i>			
2.5 Appropriate consideration of/sensitivity to the poverty dimension. For example, discussion of (i) how current policies/program affect the poor? (ii) how the proposals made by the task will affect the poor?	60	83	77
<i>Comments:</i>			
2.6 Appropriate consideration of/sensitivity to the gender dimension	NA	77	68
<i>Comments:</i>			
<i>Guidance: Were the different needs of males and females considered? [e.g. Was the data used disaggregated by sex? Were the barriers to inclusion/access discussed?] Did the report attempt to identify gender issues that might affect implementation of the policy advice or recommendation? Did the report consider the different roles played by men and women in making recommendations? Did the report consider the different impact of the recommendations on men and women</i>			
2.7 Quality of conclusions and recommendations	NA	71	74
(a) conclusions and recommendations appropriately specific and consistent with the analysis?	NA	NA	76
(b) recommendations adequately selective/prioritized?	48	55	56
(c) appropriateness of the recommendations, particularly in the light of social, political and administrative realities?	71	78	80

QAG ASSESSMENT GUIDANCE QUESTIONNAIRE

Assessment Rating
 1= Highly satisfactory
 2= Satisfactory
 3= Marginal
 4= Unsatisfactory
 NA= Not applicable

3 DIALOGUE AND DISSEMINATION

FY00	FY01	FY02
NA	87	88

General Guidance: This set of questions seeks to inquire how and to what degree the country clients, and other stakeholders such as donors, were active participants in the performance of the task under review. Effective participation and the opportunities for dialogue associated with it are recognized as hallmarks of ownership, and an important indicator of the prospects for the achievement of task objectives. Not all questions in this section may be equally relevant to any single task. Accordingly, in cases where, because of the purpose and nature of the task, selected questions are not applicable, score NA.

3.1 Attention devoted to engaging the client and key stakeholders

Comments:

a) Well-thought-out strategy for engagement

NA	NA	85
----	----	----

b) Relevant and effective counterparts

NA	68	85
----	----	----

(c) efforts at building coalitions and increasing awareness on the part of key stakeholders

NA	80	90
----	----	----

NA	NA	80
----	----	----

3.2 Effectiveness of Client participation in:

Comments:

a) defining the scope of the work

80	84	78
----	----	----

b) carrying out the analysis

NA	83	78
----	----	----

c) formulating the conclusions and recommendations

NA	75	76
----	----	----

NA	67	75
----	----	----

3.3 Quality of Bank contribution/support for client-led tasks:

Comments:

a) defining the scope of work

NA	100	100
----	-----	-----

b) organizing the task, putting the team together, selecting the consultants,

NA	82	54
----	----	----

c) knowledge sharing, providing cross-country experience

NA	100	54
----	-----	----

d) reviewing work in draft stages

NA	69	100
----	----	-----

NA	98	100
----	----	-----

Guidance: This question seeks to assess the quality of Bank support to the Client in preparing those components of the task for which the Client has had the lead or a major, substantive role. Score NA if the task is essentially Bank-led.

3.4 Quality of capacity-building efforts during task implementation

Comments:

NA	82	89
----	----	----

3.5 Appropriate consultation with other key stakeholders in the country (e.g. legislators, local government officials, civil society groups, local academics, private sector, etc. groups, local academics, private sector, etc.)?

Comments:

NA	81	82
----	----	----

3.6 Quality of partnership arrangements with other donors

Comments:

NA	78	90
----	----	----

3.7 Appropriateness of the actual/planned arrangements for dissemination

Comments:

87	87	85
----	----	----

3.8 Quality of presentation of the Bank's written products

79	88	87
----	----	----

QAG ASSESSMENT GUIDANCE QUESTIONNAIRE

Assessment Rating
 1= Highly satisfactory
 2= Satisfactory
 3= Marginal
 4= Unsatisfactory
 NA= Not applicable

FY00	FY01	FY02
88	74	86

4 LIKELY IMPACT

General Guidance: Based on the information available, including interviews with country counterparts, as appropriate, panels are asked to make judgments about expected impact of the task in relation to the development objectives of the task. The panel is not assessing actual future impact, but rather the probability of positive outcomes. Likely Impact is expected to depend on how relevant/timely the task is, how well it has been done and how receptive is the client. Panelists should seek a sense of the value-added of the task (e.g. What if the task had not been done?). The ratings should factor out substantial and specific changes in the country situation beyond the team's control. The panelists should score N/A against those issues that are not related to explicit objectives of the task. They should also be sensitive to the size and scope of the task as planned in the Concept Paper.

4.1 Efforts made towards maximizing impact

Comments:

a) clearly defined indicators of success.

b) a clear and well-structured strategy for achieving impact

(c) timeliness of output(s) in terms of the client's own cycle of policy and program review and planning

NA	71	83
NA	NA	75
NA	62	79
NA	92	89

Guidance: Does the concept paper identify clearly what are the measures of impact or results for the intended audiences?

4.2 Likely Impact on the Client

Comments:

a) helping to create ownership of change proposals among key decision-makers

b) enhancements in the area of capacity building, knowledge transfer, or institutional change

c) promoting meaningful change in government policy, strategy, expenditures, investment, programs, project design

NA	71	83
NA	69	82
NA	74	81
72	68	80

4.3 Likely impact on the Bank

Comments:

a) its strategic approach to the country

b) design/implementation of operational products

c) the knowledge-base of the Bank

NA	88	89
87	82	88
70	89	87
86	83	95

4.4 Likely impact on the wider development community:

Comments:

a) other donors (including NGOs) and aid coordination

b) development thinkers

Comments:

80	78	97
NA	80	94
NA	63	95

Guidance: This is not a 'necessary' requirement of high impact. Score NA where objectives of the task did not anticipate spin-offs to the wider development community.

QAG ASSESSMENT GUIDANCE QUESTIONNAIRE

Assessment Rating
 1= Highly satisfactory
 2= Satisfactory
 3= Marginal
 4= Unsatisfactory
 NA= Not applicable

	FY00	FY01	FY02
5 BANK PROCESSES	68	66	83

INPUTS

5.1 Elapsed time (months)			
Activity initiation – Discussion Draft	9	11	13
Discussion Draft - Client Delivery.....	4	5	3
5.2 Preparation/Processing Cost (\$000)			
Regular Budget (BB).....	206	185	205
Trust Funds.....	45	70	30
Total Funding	251	255	235

MANAGERIAL ATTENTION

5.3 How reasonable was the time taken to complete the task?	82	78	95
5.4 How adequate was the total funding in view of the objectives and scope of the task? Funding for dissemination & follow-up?	65	67	85

Guidance: Bear in mind the complexity, country context, previous work available. Do not overlook TF or other donor resources. Was dissemination sufficiently funded?

Check if:

5.5 Skill-mix of the task team? How comprehensive? Cross-sectoral skills?	82	82	88
5.6 Overall efficiency of resource use (both Bank budget and Trust Funds, where used)	84	88	91

Guidance: Did the managers select the right leadership? Did the manager or TTL assemble the right team in terms of country knowledge, experience, skills?

Comments:

a) Bank Budget

NA	NA	91
----	----	----

b) trust funds component only

NA	NA	93
----	----	----

5.7 Attention to Quality at Entry

Comments:

a) by Sector Management

69	64	85
----	----	----

b) by Country Management

67	73	83
----	----	----

Guidance: Extent of shared understanding between managers and the team on the objectives, methodology, key milestones for progress and indicators of success, etc; an appropriate understanding on roles of core participants at the time of launching the task - through a Concept Paper or otherwise. Sector management here means the individual with a direct supervisory/advisory relationship with TTL in the performance of the task.

5.8 Attention to implementation/dissemination of the task.	60	76	83
---	----	----	----

Comments:

a) by Sector Management

69	80	82
----	----	----

b) by Country Management

77	76	83
----	----	----

Guidance: Monitoring of quality/milestones, promotion of cross-fertilization, best practices, etc. Did the management provide appropriate guidance and mentoring? Were adjustments made and mid-course corrections taken in a timely manner? Were comments on drafts appropriately considered? Did management support and/or participate in the dissemination and dialogue? Typically such attention is at key points rather than continuous and is provided in relation to the size, complexity of the task and the experience, skill mix of the task team.

OTHER

5.9 Quality enhancement contributions of the Network/Sector anchor(s).

Comments:

a) Quality of inputs

b) Impact on final output

Guidance: Score NA if the team did not seek any support from the anchor(s). Use of available supply of knowledge produced by the anchors is covered in question 2.4.

5.10 Quality enhancement contributions of the peer review process

Comments

a) Quality of inputs

b) Impact on final output

5.11 Quality enhancement contributions of a QER or equivalent

Comments:

a) Quality of inputs

b) Impact on final output

5.12 Quality of the Activity Completion Summary?

Comments:

5.13 Quality of information on the task in SAP

Comments:

Guidance: Completeness/accuracy of budget information, task description, task objectives, milestones etc

FY00	FY01	FY02
------	------	------

NA	87	96
----	----	----

69	89	96
----	----	----

58	85	82
----	----	----

NA	63	76
----	----	----

95	86	83
----	----	----

71	62	70
----	----	----

NA	100	100
----	-----	-----

NA	100	100
----	-----	-----

NA	100	100
----	-----	-----

67	44	52
----	----	----

NA	67	70
----	----	----

ANNEX 6

TABLE 1. TTL SURVEY

	<u>Satisfactory (%)</u>	
	FY02 (n=40)	FY01 (n=22)
Performance of the QAG panel		
a) Panel's skills mix in relation to the task's subject and coverage	83	81
b) Panel's familiarity with country specific issues	65	n/a
c) Panel's preparation for the interview	88	81
d) Extent to which the panel focused on the relevant issues	83	67
e) Panel's listening to the task team's point of view	80	n/a
f) Appropriateness of the tone and atmosphere during interview	98	n/a
ESW5 Process		
a) Transparency of the process	85	n/a
b) Role of the moderator	92	n/a
c) Logistics of the process	88	n/a
d) Time taken to share the draft assessment with the team	85	n/a
While QAG review is essentially an accountability exercise, how Would you rate it as a learning experience?	65	72
Overall QAG review process	74	71
Average time (in hrs) taken for:		
Document preparation	16	18
Panel interview	4	4

ANNEX 6

TABLE 2. PANELIST SURVEY

	<u>Satisfactory (%)</u> (n=83)
Assessment Panel	
a) Panel's skills-mix in relation to the task's type and complexity	97
b) Panel's familiarity with country specific issues	84
Assessment Approach	
a) Usefulness of the TTL inputs on 'context' (Part A of the questionnaire) towards assessment	92
b) Appropriateness of the Guidance Questionnaire to the assessment	97
c) Usefulness of the client feedback to the assessment	64
d) The other panelist were well prepared for the assessment	96
e) My views were taken into account when rating the task	97
Role of Moderator	
a) Role in setting the ground rules and the tone of the discussion	97
b) Guidance provided in interpreting the questionnaire	97
c) Guidance provided towards benchmarking the ratings vis-à-vis similar tasks	92
Logistics	
a) Task background documentation - timely and appropriate?	92
b) Quality of QAG Administrative Support	97
c) Adequacy of the time provided to conduct the assessments	92
Overall Performance	
Usefulness of the ESW5 process overall for enhancing accountability in the Bank?	92
Usefulness of the ESW5 process to identify systemic issues	92
Cost effectiveness of the ESW5 process.	98
Average panelist time per assessment	20.7 hrs