



# **Quality of Supervision in FY05-06 (QSA7) A QAG Assessment**

**June 14, 2007**



QUALITY ASSURANCE GROUP

# TABLE OF CONTENTS

## EXECUTIVE SUMMARY

Slides i-vi

## I. OBJECTIVES AND METHODOLOGY

- Objectives
- Methodology
- Sample and Robustness of Results
- Do QSA Ratings Matter

Slide 1  
Slides 2-3  
Slides 4-5  
Slide 6

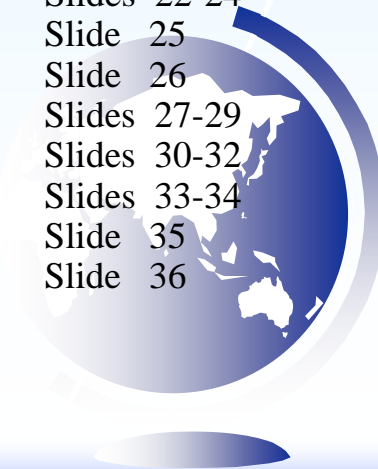
## II. SUPERVISION CONTEXT

Slides 7-8

## III. ASSESSMENT FINDINGS

- Bankwide Results
- Results by Quality Dimension
- Results by Source of Funds
- Results by TTL/Staff Location
- Operations Rated Highly Satisfactory
- Operations with Unsatisfactory Supervision
- Operations Rated Moderately Satisfactory
- Results by Region
- Results by Network
- Results by Sector Board
- Results by LICUS/Non-LICUS
- Results by Loan Size, Project Age and At Risk Category

Slides 9-12  
Slides 13-17  
Slides 18-20  
Slide 21  
Slides 22-24  
Slide 25  
Slide 26  
Slides 27-29  
Slides 30-32  
Slides 33-34  
Slide 35  
Slide 36



# TABLE OF CONTENTS

<b>IV.</b>	<b>PORTFOLIO MANAGEMENT INDICATORS</b>	Slides 37-39
<b>V.</b>	<b>RECOMMENDATIONS</b>	Slides 40-41
<b>VI.</b>	<b>PREVIOUS QAG RECOMMENDATIONS</b>	
	▪ Project-level Assessments	Slide 42
	▪ Status of Follow-up on QSA6	Slides 43-45
<b>VII.</b>	<b>FEEDBACK ON QSA7 PROCESS</b>	Slides 46-47



# EXECUTIVE SUMMARY

## Background

- QSA7 is the seventh of a series of assessments of quality of supervision of Bank lending operations aimed at increasing accountability and promoting systemic improvements. As in previous assessments, QSA7 focuses on Bank rather than project performance.
- QSA7 covers 130 randomly selected operations approved prior to FY06 and active as of July 1, 2006, representing 11% by number and dollar amount of the universe.
- The sample size is set to permit robust results at Bank-wide level (confidence interval: 95%; margin of error: +/- 5%) and for most regions and networks.
- This is the second QSA utilizing a 6-point scale ranging from Highly Satisfactory to Highly Unsatisfactory. The six point scale enables the identification of areas in need of further improvement or missed opportunities that would otherwise be lost in a pass-fail system. For accountability purposes, the top three ratings are considered Satisfactory and the other three, Unsatisfactory.



## Key Findings

- The share of projects with unsatisfactory supervision went down from about 10% in the last three assessments to 5% now suggesting that the instances of supervision with serious flaws have become rare. This is a major achievement compared to the situation 10 years ago when supervision for one-third of the sampled projects was found to have major deficiencies.
- The gains in supervision quality are broad based with all regions and networks meeting or exceeding the agreed target of 90% Satisfactory level. Particularly noteworthy in this respect is the improved performance by AFR, MNA and SAR.
- Performance of three of the four quality dimensions (Focus on Development Effectiveness, Supervision of Fiduciary/Safeguard Aspects, and Adequacy of Supervision Inputs and Processes) also meets or exceeds the 90% level, the sole exception being Candor and Quality of Supervision Reporting.

...*Cont'd*

## Key Findings

- Notwithstanding the commendable strides made in reducing unsatisfactory supervision, QSA7 results also point to significant scope for further improvement through focus on missed opportunities.
  - ✓ Only half the projects in QSA7 are rated as fully Satisfactory or better compared to 65% in QSA6 due to significant increase in the share of the Moderately Satisfactory category.
  - ✓ Performance reporting of QSA7 projects understates the riskiness of the portfolio in the Bank system. The Panels identify twice as many problem projects and three times as many risk flags in both FY05 and FY06 than are reported in the ISRs. This indicates that more projects need special attention and possibly additional resources.
  - ✓ Focus on Development Effectiveness is a key concern in QSA7 and is rated lower than in QSA6 in terms of S+. Panels note satisfactory focus on management of inputs and outputs, less so on outcomes. Weaknesses in Quality and Timeliness of Data to Support KPIs constrain the ability of teams and management to identify and address in a timely manner the threats to DOs.

*...Cont'd*

## Key Findings

- Further improving performance would require addressing the following aspects:
  - ✓ QSA7 results indicate that management attention could be improved further, signaling a need to review incentives and workload of sector/country managers to ensure staff mentoring, upgrade the quality of reporting, and deal with supervision issues in a timely manner.
  - ✓ Quality of supervision skills mix also shows a decline from QSA6, and requires more attention specifically with respect to: (a) better support to decentralized staff in terms of complementary global and specialized skills; (b) stronger mix in the social and environment areas; and (c) better integration of decentralized fiduciary staff with supervision teams.
  - ✓ About 20% of the sample are rated as having inadequate budgets in part linked to understatement of risks. Greater candor would help in enabling the tailoring of budget allocations to project needs.



# Recommendations

## *Senior Management*

- ✓ Strengthen staff and managerial incentives and accountabilities for candor and quality of supervision reporting.

## *Regions*

- ✓ Improve quality and timeliness of Mid-term Reviews and encourage restructuring of problem projects as early as possible.
- ✓ Ensure that supervision budget allocations are tailored to project needs to maximize impact.
- ✓ Ensure that procurement and financial management staff are better integrated with supervision teams.

...Cont'd



# Recommendations

## *Networks/Sector Boards*

- ✓ Provide guidance and training to task teams to improve the results framework of projects and promote dissemination of best practice examples across regions.

## *QAG*

- ✓ Explore options for improving the value added from future QSAs in light of progress made to date in supervision quality.



# I. OBJECTIVES AND METHODOLOGY

## Objectives

- **Promote excellence in Bank performance by:**
  - **Increased accountability** - Provide *credible feedback in real time* to staff and managers on the quality of their supervision work;
  - **Enhanced learning** - *Identify lessons of experience and use them to catalyze changes* in policies, programs and procedures;
  - **Improved Supervision of LICUS projects** – *Identify areas of improvement* in the supervision of LICUS projects; and
  - **Improved Quality of Performance Indicators** – *Provide feedback on the realism of performance indicators* as reported in the ISRs.



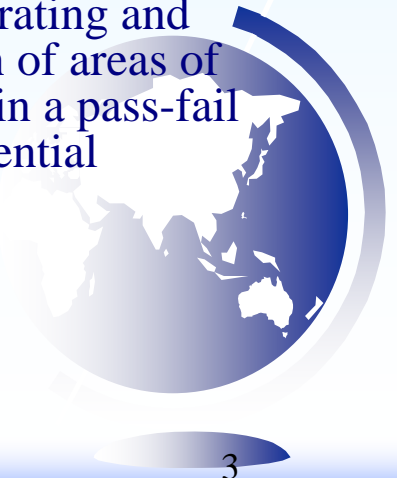
## Methodology

- The methodology and guidance questionnaires were essentially the same as in previous assessments, with sharpening of questions relating to Results and ISR.
- Panelists used a 6-point rating scale (*Highly Satisfactory, Satisfactory, Moderately Satisfactory, Moderately Unsatisfactory, Unsatisfactory* and *Highly Unsatisfactory*) similar to the scale used by IEG.
- In terms of accountability, the first three ratings – Highly Satisfactory, Satisfactory, and Moderately Satisfactory – are considered *Satisfactory* performance while the other three ratings constitute *Unsatisfactory* performance.
- A Stage 2 assessment was available to all operations rated *less than Moderately Satisfactory* in the first review and was used in eight cases leading to an upgrade in two operations.
- To promote learning, 60 Bank staff participated as observers.

...Cont'd

# Methodology

- 95 Staff and 48 Consultants (mostly retired Bank Staff) participated as Panelists. 14 Moderators with extensive experience in QAG assessments ensured consistent application of the methodology and ratings. In addition, the QAG Director reviewed all assessments to further ensure consistency across the board and over time (i.e., in comparison with previous QSAs).
- While the methodology provides guidelines on how to determine ratings, there is an element of subjectivity in the process. This is mitigated by the design of the questionnaire which helps ensure internal consistency, as well as by different levels of review including by outside parties.
- The shift to a 6 point scale starting with QSA6 makes the rating process more of a continuum, with some blurring of distinction between one rating and another. However, the six point scale enables the identification of areas of concern and missed opportunities that would otherwise be lost in a pass-fail system, and should lead to more specific actions to address potential emerging weaknesses in the supervision process.



# Sample and Robustness of Results

- **The QSA7 sample includes 130 operations out of 1,193 approved prior to FY06 and still active on July 1, 2006.** It covers 11% by number as well as 11% by dollar amount of this universe. By source of funding the sample includes: 74 IDA, 41 IBRD, 10 GEF, 2 Montreal Protocol and 3 SPF. By lending instrument, the sample includes: 87 sector investment loans, 15 technical assistance loans, 16 APLs, 4 ERLs, 3 SIMs, 2 LILs, 1 FIL and 2 DPLs.
- **The sample was stratified first by Region and then LICUS/non-LICUS.** Excluded from the sample were operations approved in FY06 – the time for implementation and supervision was too short to permit valid conclusions.
- **Sampling fractions** were chosen for each Region based on *a priori* probability indicated by QSA5-6 results:
  - Regions with higher proportion of overall *Satisfactory* ratings in QSA5-6 and with a large number of projects were sampled at a lower rate in QSA7;
  - Minimum sample size for any Region was set at 15 operations;
  - Sample sizes by Regions were: AFR - 35; EAP - 21; ECA - 27; LCR - 17; MNA - 15; and SAR – 15; and
  - Sample sizes by Network were: ESSD – 39; INF – 39; HDN – 29; PREM – 12; PSD – 7; FSE – 4.

...Cont'd

## Sample and Robustness of Results

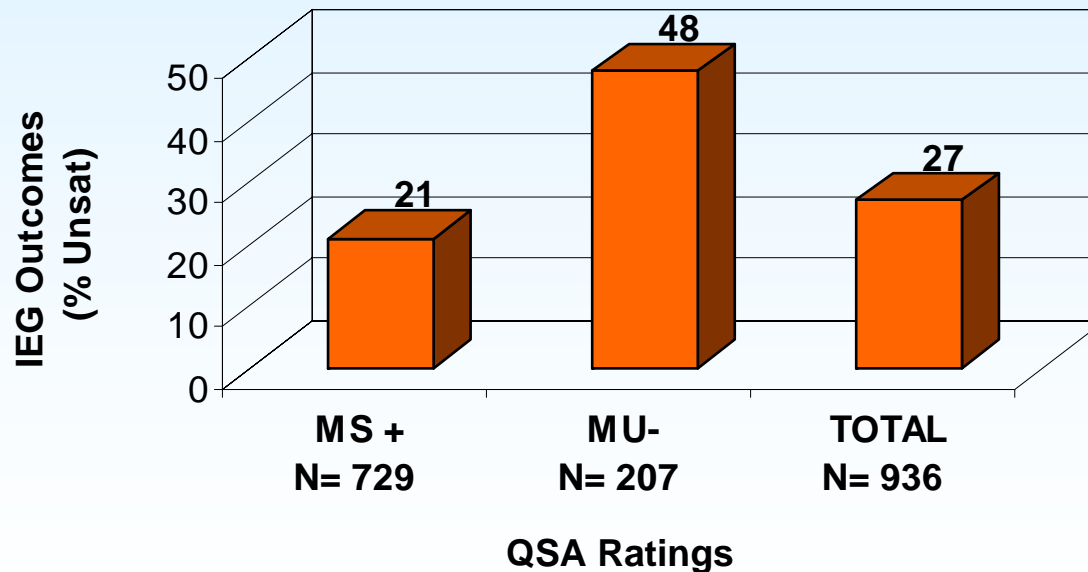
- **Weighting:** Each operation in the sample has been assigned a weight which is the inverse of the ratio of the sample size in the respective stratum for Region and LICUS/non-LICUS to the portion of the universe from which it is drawn.
- **Results in this report are presented on a weighted basis** to adjust for over/under-representation in the sample because of stratification and sample design.
- **QSA7 results are statistically robust for the following:**
  - Bank-wide at 95% +/- 5%;
  - Regions at 90% +/- 10% except for MNA and SAR at 90% +/- 15%;
  - ESSD, INF, and HD Networks at 90% +/- 10%;
  - LICUS at 90% +/- 6%; and
  - Combined QSA6-7 Sector Board results at 90% +/- 10% for RDV, ED, ENV, PSG, EMT, TR, and WS.



## Do QSA Ratings Matter?

- IEG evaluations of projects which were assessed by QAG in various QSAs, and which have since exited the portfolio, show a clear and strong association between Satisfactory ratings in QAG supervision assessments and Satisfactory IEG outcome ratings. Using data as of May 2007, the IEG ratings of 936 QSA operations show the following results:

*Figure 1: IEG Outcomes vs. QSA Ratings*



## II. SUPERVISION CONTEXT

### Development Objectives

- 78% of the operations in the sample are rated Moderately Likely or better (ML+) to achieve the Development Objectives, about the same as the IEG rating for achieving development outcomes of operations exiting the portfolio during FY04-06.
- There is a large gap between the ML+ and L+ ratings of DOs. Less than half of the operations were rated L+ in terms of achieving DOs. Improving the results framework at entry and appropriate and timely restructuring of projects during supervision would help ensure greater achievement of realistic objectives.
- Institutional Development and Capacity Building is by far the most cited DO across regions and sectors. Panels rate the likelihood of achieving this DO at 83% ML+.



# Quality at Entry

- Design at entry is rated at 80% MS+. This is lower than comparable results in QSA6 as well as in QEA7 where overall quality at entry was rated 92% MS+. As important, the S+ rating for design at entry in QSA7 was only 33%.\*
- The main issues at entry are weaknesses in the following areas:
  - Quality of the results framework and M&E;
  - Readiness of implementation at approval; and
  - Extent to which the problems encountered during implementation have been identified at entry (risk assessment).

\* The QSA results should be viewed with caution when comparing with QEA given that quality at entry reviews by QSA panelists are not as intensive as those in QEA, though QSA has the benefit of hindsight. In addition, there may be an incentive for TTLs to magnify quality at entry issues when explaining gaps or weaknesses in supervision performance.



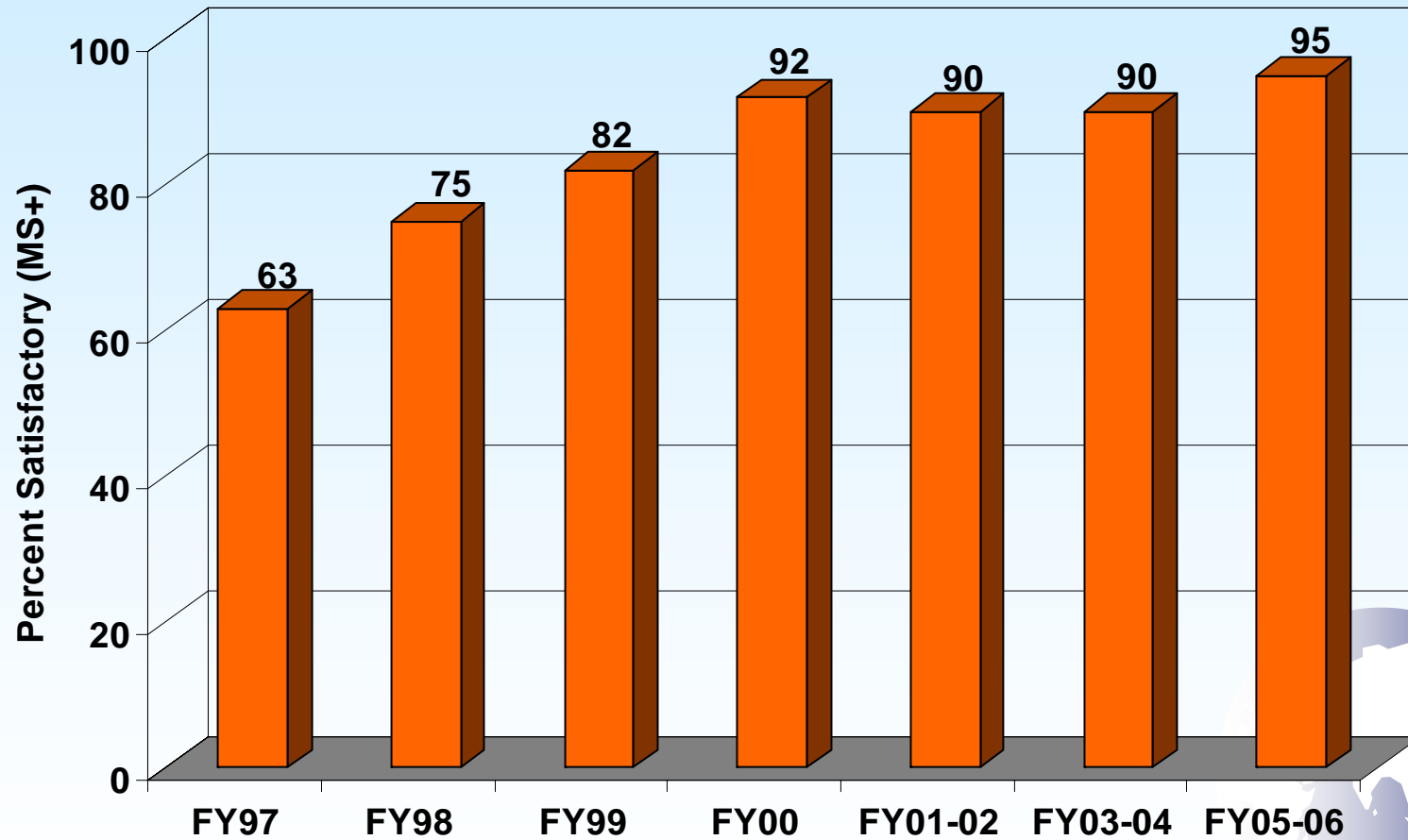
# III. ASSESSMENT FINDINGS

## Bank-wide Results

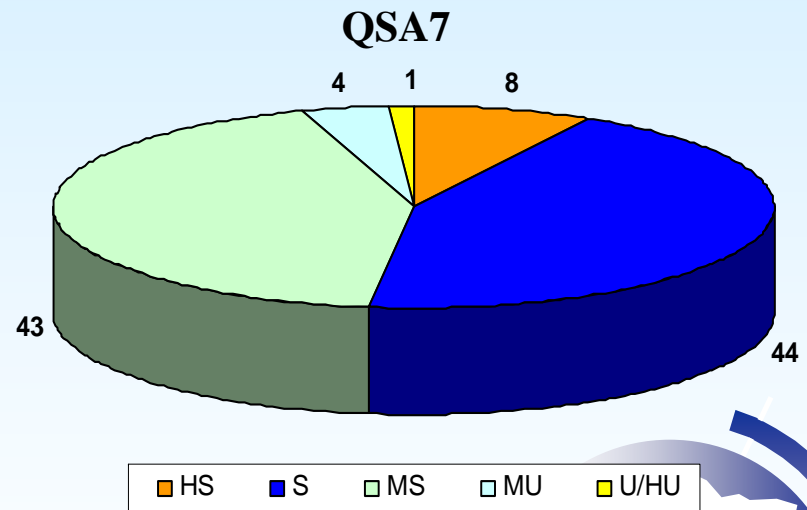
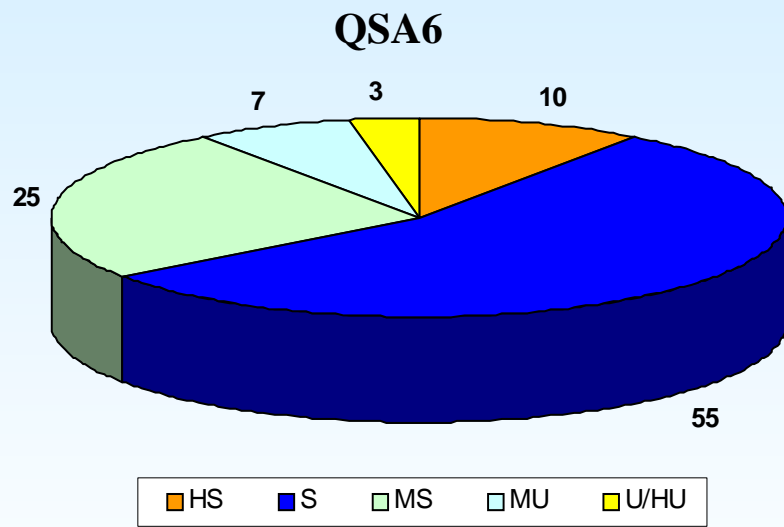
- In terms of MS+, QSA7 overall results show continuity of a positive trend since FY00 (Figure 2) and an improvement over QSA6 in all dimensions.
- All except one quality dimension score 95% or above in terms of MS+ (Figure 4). Candor and Quality of ISR is rated at 85% MS+.
- All of the main aspects of supervision quality are at 85% or better in terms of MS+ with the exception of: (a) performance monitoring; (b) adequacy of supervision budget; (c) accuracy of ratings or project components and discretionary flags; and (d) quality and timeliness of data to support the key performance indicators.
- However, the 52% S+ performance in QSA7, which is lower than in the 65% figure in QSA6 (Figure 3), indicates room for improvement as well as missed opportunities.



*Figure 2: Overall Assessment of Supervision Performance  
(FY97-FY06)*



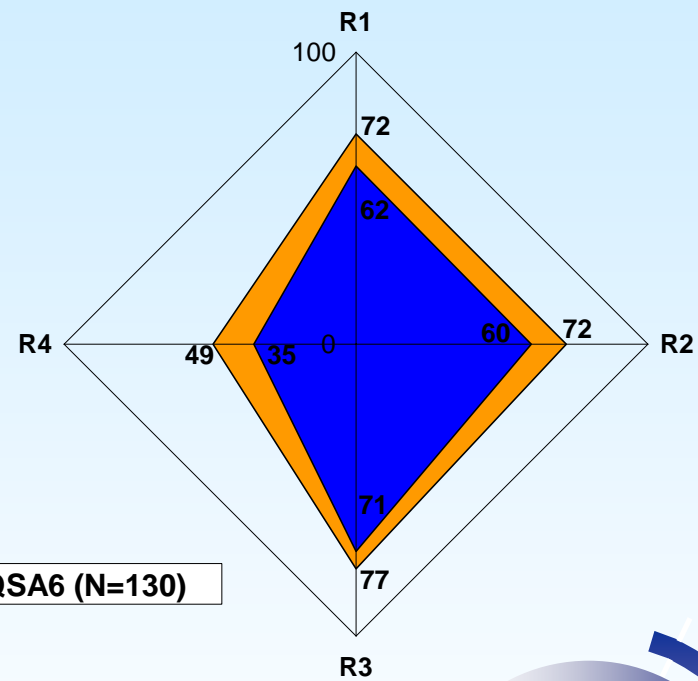
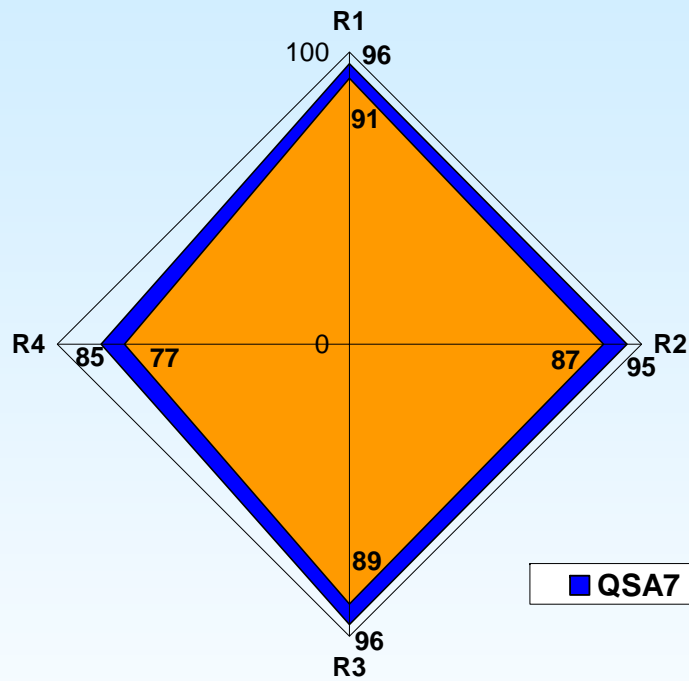
*Figure 3: QSA6 and QSA7 Bankwide Ratings (%)*



**Figure 4: Results by Quality Dimension**

**(% Moderately Satisfactory or Better)**

**(% Satisfactory or Better)**



■ QSA7 (N=130)    ■ QSA6 (N=130)

- R1: Focus on Dev. Impact
- R2: Supervision of Fiduciary and Safeguard Aspects
- R3: Adequacy of Spvn. of Inputs and Processes
- R4: Candor and Quality of Proj. Performance Reporting



# Results by Quality Dimension

## Focus on Development Effectiveness (96% MS+ and 62% S+)

- This dimension shows improvement from 91% to 96% in terms of MS+ but a deterioration from 72% to 62% in terms of S+ when compared to QSA6 (Figure 4).
- Strong aspects include:
  - Identification and assessment of problems; and
  - Focus on sustainability.
- Improving performance under this dimension calls for:
  - Focusing more strongly on outcomes, including further strengthening of M&E; and
  - Improving the timeliness and quality of Mid-Term Reviews as well as restructuring projects in a timely manner.



## **Fiduciary and Safeguard Aspects (95% MS+ and 60% S+)**

- There is an improvement in this dimension compared to QSA6 from 87% to 95% in terms of MS+, but a deterioration in terms of S+ from 72% to 60% (Figure 4).
- **Procurement** performance at 97% MS+ continues to be high reflecting Satisfactory supervision in all aspects, notably timely identification of and response to procurement issues arising during project implementation. However, at 78% S+ there is scope for further improvement specifically in the following areas: improving risk assessment and mitigation; effective utilization and improved quality of procurement plans; and consistent means of tracking post reviews which play an increasingly important role in enhancing anti-corruption focus.
- **Financial Management** performance at 94% MS+ sustains the high level of performance reflecting Satisfactory supervision in all aspects, including providing support to Borrower. However, at 69% S+, there is scope for further improvement primarily in record keeping aspects. Other recommended actions to further strengthen supervision include: improved guidance in the FM Manual; more consistent application of supervision principles across regions; provision of tools and guidance to staff for enhanced anti-corruption focus.

*...Cont'd*

- **Legal Aspects** at 97% MS+ and 89% S+ reflects high performance in ensuring that legal requirements are met.
- **Environmental Aspects** at 90% MS+ sustains the level of performance since FY00 but the number of non-compliance cases doubled from 3 in QSA6 to 7 in QSA7, all of them concerning Environmental Assessment (OP 4.01). Monitoring of EMP or Variant, while at 90% MS+, shows a decline from 95% S+ in QSA6 to 79% S+ in QSA7. Areas of improvement are in ensuring availability of skilled staff and upgrading the implementation quality of EMPs.
- **Social Aspects** at 87% MS+ is at about the same level of performance as in QSA6 reflecting high level of performance in ensuring compliance with safeguards, of which there are only 4 substantive and 3 procedural cases of non-compliance, compared to 13 cases in QSA6. However, with 59% S+ there are challenges especially in the following areas: identifying and measuring social development outcomes, follow-up on gender and inclusion issues identified at entry, supervision of social and political risks, and M&E of poverty and social outcomes. Greater attention to these areas – both at entry and during supervision – would help improve performance in social aspects beyond compliance with safeguards.



## Adequacy of Supervision Inputs and Processes (96% MS+ and 71% S+)

- This dimension shows a slight improvement from 89% to 95% in terms of MS+ and slight deterioration from 77% to 71% in terms of S+ (Figure 4).
- All aspects within this dimension are at better than 90% MS+ with the exception of Adequacy of Supervision Budget at 78%. About 20% of projects were considered by Panelists to have too little budget and 4% too much.
- Impact of Decentralization on Supervision Quality is rated at 99% MS+. Panelists note, however, that there is scope for improving integration of fiduciary specialists in the supervision process.
- To improve performance towards Fully Satisfactory, the following areas need attention:
  - Quality of Supervision Skills Mix shows a decline from 77% to 62 % in terms of S+; and
  - Adequacy and Speed of Management Attention and Actions shows a decline from 75% to 61% in terms of S+.



## **Candor and Quality of ISR (85% MS+ and 35% S+)**

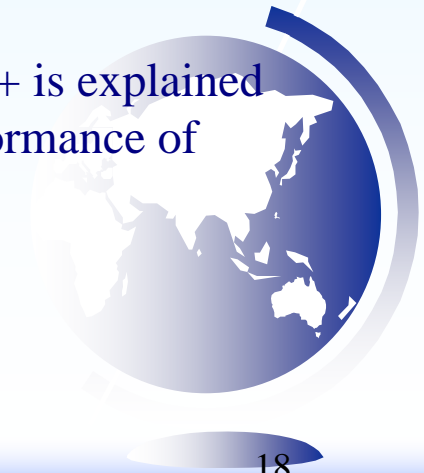
- This dimension shows an improvement from 77% to 85% in terms of MS+, but a decline from 49% to 35% in terms of S+ (Figure 4). In terms of MS+, Candor and Quality of ISR is rated lower by ten percentage points or more than other quality dimensions.
- Extent to Which Ratings for DO, IP and Risk Ratings Reflect Actual Conditions improved from 77% to 88% in terms of MS+ but declined from 67% to 48% in S+. Panelists estimate that at end-FY06, 29% of the sample have problems in DO, IP, or both, compared to ISR estimates of 12%.
- Quality and Timeliness of Data to Support Key Performance Indicators is rated at 79% MS+ and 42% S+ (there is no comparable data in QSA6). This performance has its roots in weaknesses in the results framework at entry and contributes to poor identification and assessment of threats to DO and delayed implementation of corrective actions. Improving the results framework at entry, including ensuring timeliness and quality of baseline data, would ensure better performance in this aspect of supervision.



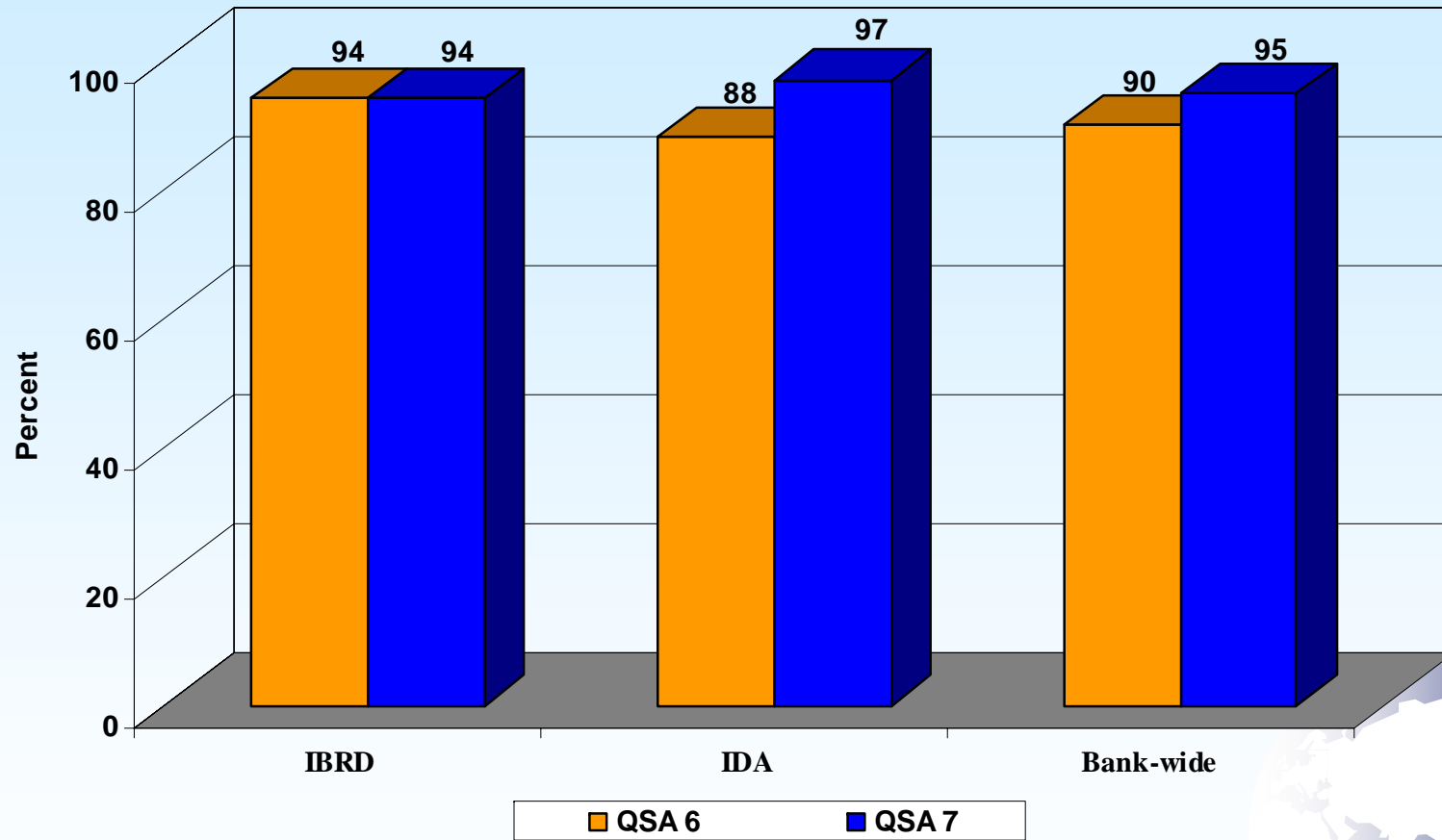
# Results by Source of Funds

## IBRD/IDA

- Performance of IDA projects in QSA7 improved from 88% to 97% in terms of MS+ and its performance is now at par with that of IBRD projects (Figure 5).
- In terms of S+, IDA projects scored better at 56% S+ compared to 41% S+ for IBRD funded projects. Performance of IBRD projects declined from 69% S+ to 41% S+ (Figure 6).
- The difference between IDA and IBRD projects in terms of S+ is explained by improved performance of AFR and the decline in the performance of regions with large IBRD portfolio.

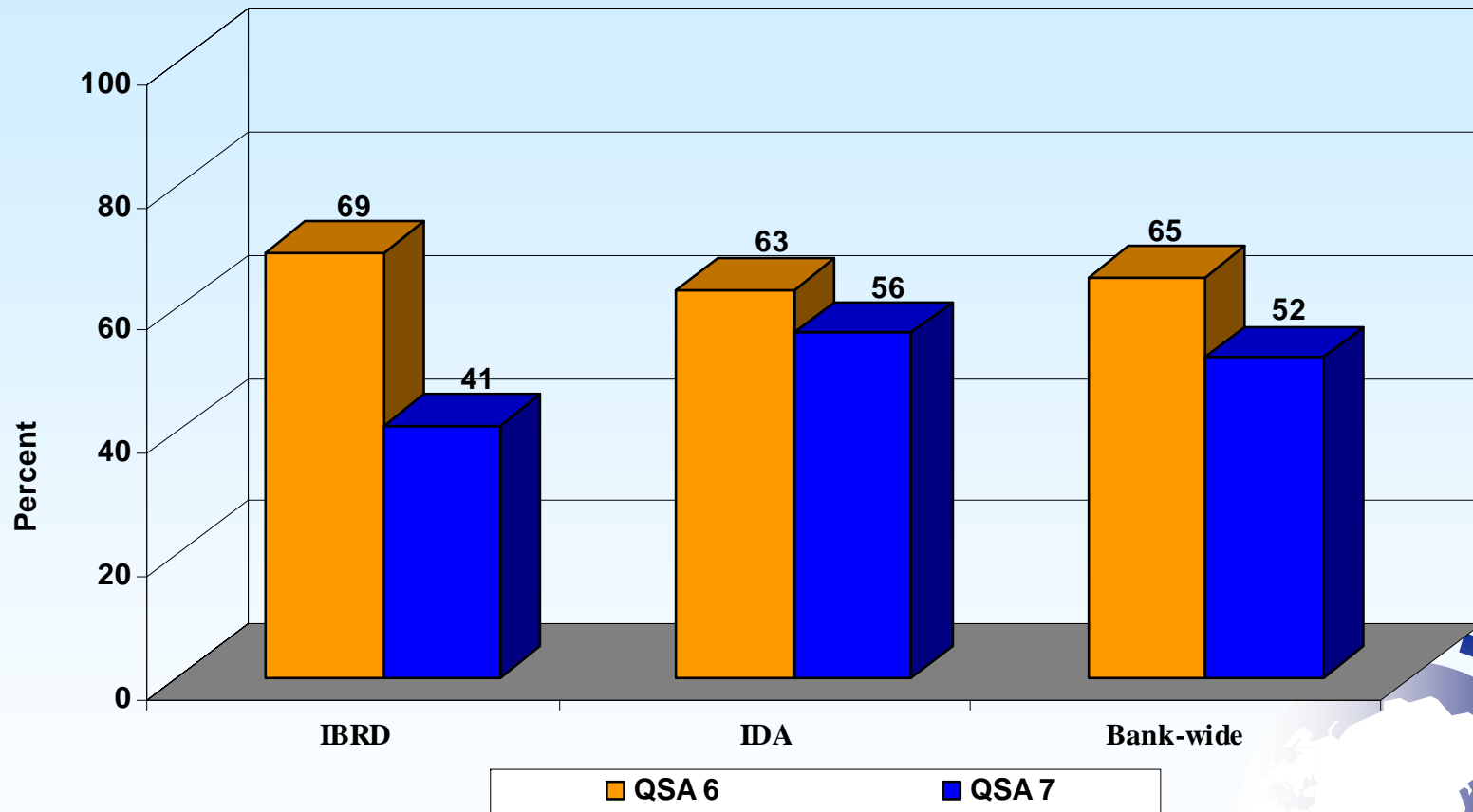


**Figure 5: Results by Source of Funds  
(% Moderately Satisfactory or Better)**



Notes: 1. IBRD includes 4 IBRD/IDA and 1 IBRD/GEF blends.  
2. IDA includes 4 IDA/GEF blends.

**Figure 6: Results by Source of Funds  
(% Satisfactory or Better)**



Notes: 1. IBRD includes 4 IBRD/IDA and 1 IBRD/GEF blends.  
2. IDA includes 4 IDA/GEF blends.

# Results by TTL/Staff Location

## TTL/Staff Location

- There has been a positive impact of decentralization, including that of TTL, sector, procurement, or financial management staff, on supervision quality.
- Panelists indicate that more can be done to complement local presence with appropriate use of HQ staff with global experience and to improve the effectiveness of sector management oversight.
- In terms of S+, HQ-based supervision tend to do better in Focus on Development Effectiveness while field-based supervision do better in Supervision of Fiduciary/Safeguard Aspects. In terms of MS+, there are no significant differences of performance by staff location.



## Operations Rated Highly Satisfactory

- 10 Operations (8 % of sample) are rated Highly Satisfactory (Table 1).
- Operations rated Highly Satisfactory are characterized by:
  - Well-defined supervision strategy based on clear project objectives, strong focus on outcomes, and effective M&E;
  - A stable team with proper skills mix which combines global experience with local knowledge and presence;
  - Strong support from both Country and Sector Management with issues brought in a timely manner to the attention of key decision makers;
  - Early identification of risks to DO and IP and addressing them strategically, including through timely restructuring and Mid-Term Reviews; and
  - Strong focus on ensuring Borrowers' ownership.



***Table 1: Operations Rated Highly Satisfactory***

<b>Region</b>	<b>Country</b>	<b>Sector</b>	<b>Project Name</b>
AFR	Guinea	Health, Nutrition, Population	Multi-Secal AIDS SIL
AFR	Mali	Financial Sector	Finance Sector Development
AFR	Sierra Leone	Public Sector Governance	Institutional Reform & Cap Bldg
EAP	China	Energy and Mining	Shanghai Waigaoqiao
ECA	Bosnia-Herzegovina	Private Sector Development	Business Enabling Env SAC
ECA	Bulgaria	Energy and Mining	Energy Efficiency (GEF)
ECA	Georgia	Private Sector Development	Enterprise Rehabilitation
ECA	Turkey	Environment	ODS Phaseout 2
MNA	Yemen	Education	Basic Education Expansion Project
SAR	India	Rural Sector	UP Sodic Lands II



***Box 1: Good Practice Example: Sierra Leone Institution Reform  
and Capacity Building TAL***

- ***A difficult country context for supervision.***
  - ✓ Complex post-conflict environment; and
  - ✓ Opposition by sectoral ministries and traditional chiefs to reform.
- ***What did supervision achieve?***
  - ✓ Focus on building support to reforms by facilitating workshops that led to an acceptable clarification of roles (the Makeni Agreement).
- ***What was good about Bank supervision?***
  - ✓ A good mix from HQ (TTL), in-country, and neighboring countries (fiduciary specialists);
  - ✓ Effective collaboration with main donors DFID and UNDP ensuring the dovetailing of their local governance operations with the Bank's; and
  - ✓ Effective use of the CPPR and country dialogue to align the Bank projects with the decentralized framework for service delivery.

# Operations with Unsatisfactory Supervision

- 8 Operations (5% of sample) are rated Moderately Unsatisfactory or worse.
- Half of the operations in this category are LICUS.
- Common characteristics of Unsatisfactory Supervision are:
  - Poor design at entry with the teams unable to correct weaknesses in project design in a timely manner;
  - Lack of attention to development effectiveness with unclear project objectives and weak M&E;
  - Inadequate management attention and support;
  - Lack of candor and quality of ISR including in ratings of project performance; and
  - Inadequate skills mix.



# Operations Rated Moderately Satisfactory

- 58 Operations (43% of sample) are rated Moderately Satisfactory and hence of acceptable quality. However, improving overall quality of supervision requires more attention to the MS category which is sizeable and which increased from QSA6.
- These operations score very high in terms of MS+ in all dimensions except for Candor and Quality of ISR at 79% MS+.
- The areas of improvement for this category are:
  - Focus on Development Effectiveness (28% S+): Performance Monitoring (26% S+) and Actions Taken and Follow-up (42% S+);
  - Candor and Quality of ISR (9% S+): Accuracy of Ratings of Project Components and Discretionary Flags (16%) and Quality and Timeliness of Data to Support KPIs (23% S+);
  - Adequacy of Supervision Inputs and Processes (46%): Management Inputs at (43% S+) and Supervision Skill Mix at (41% S+); and
  - Supervision of Fiduciary/Safeguard Aspects (36% S+): Environmental Aspects (48%) and Social Aspects (40% S+).

\* Ratings are for the MS cluster of projects and not for total sample.

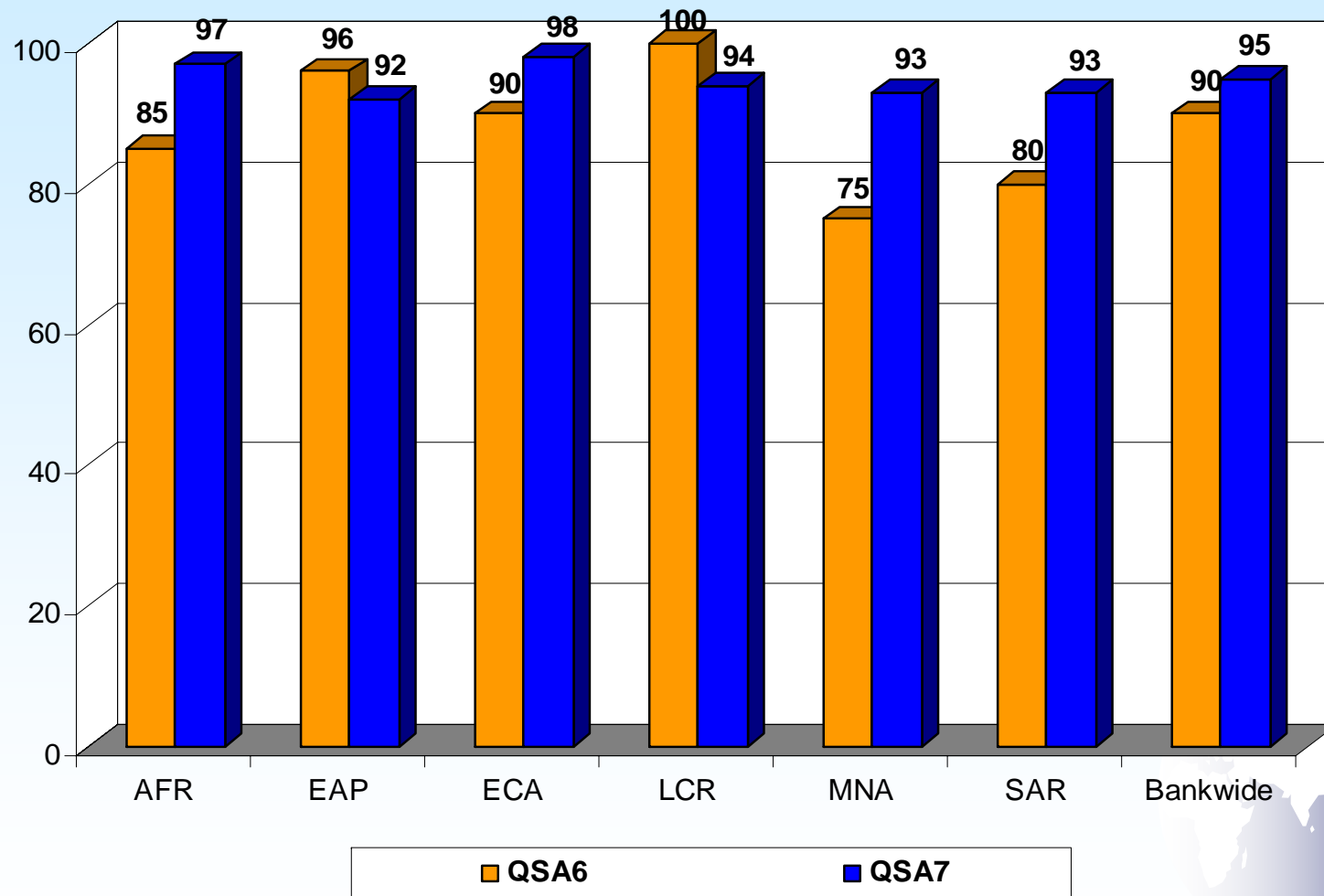


## Results by Region

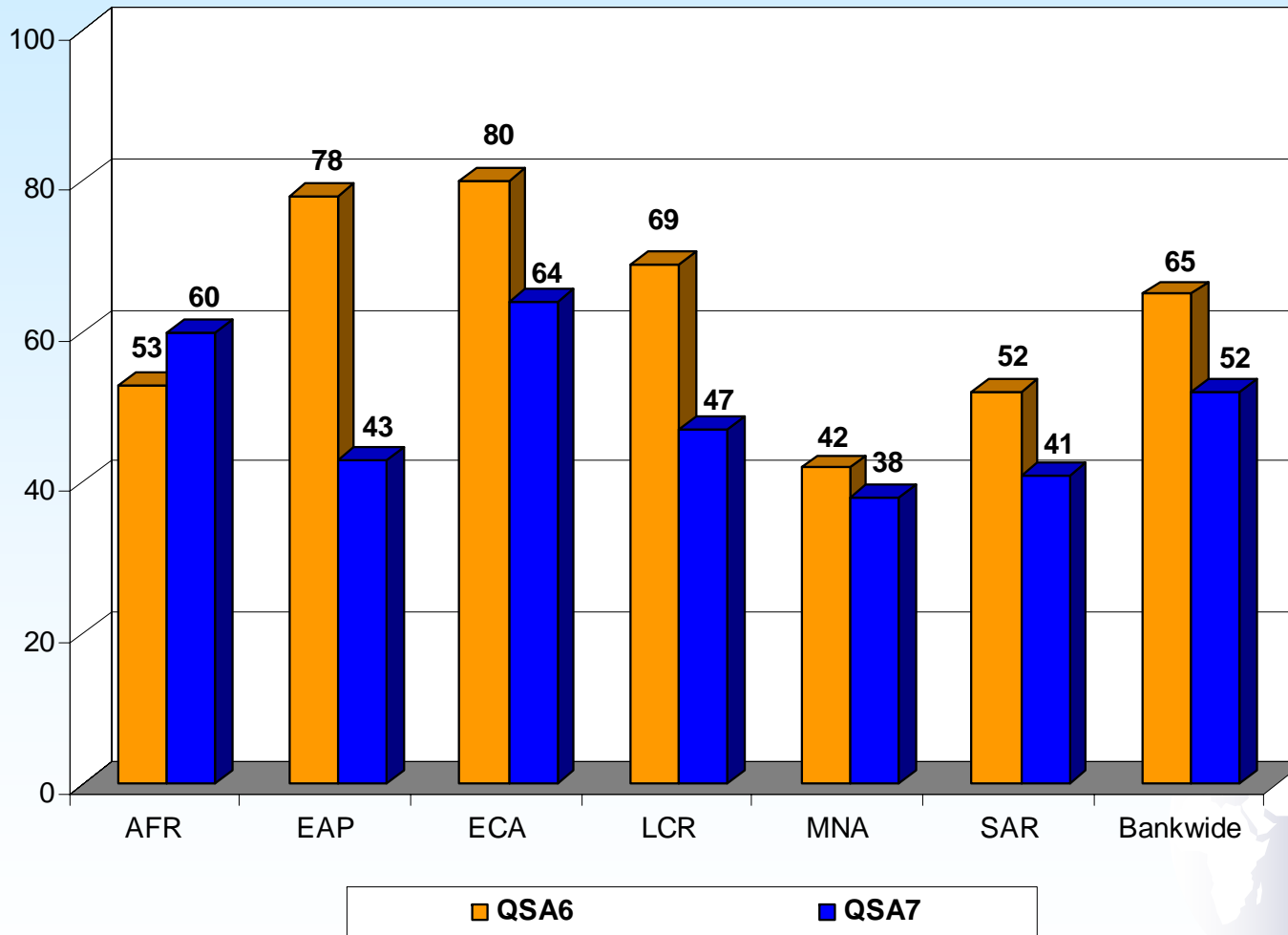
- There is convergence among regions in the MS+ category (Figure 7) with no statistically significant differences in overall performance.
- AFR, MNA, and SAR show improved performance (more than 10 percentage points from QSA6) in terms of MS+. ECA, EAP and LCR continue to be at the same high level of performance in terms of MS+.
- There continues to be a wide variation in terms of S+ (Figure 8) with ECA and AFR showing the best performance among regions. Three regions (ECA, EAP, and LCR) show statistically significant declines in terms of S+.



**Figure 7: Results by Region**  
**(% Moderately Satisfactory or Better)**



**Figure 8: Results by Region  
(% Satisfactory or Better)**

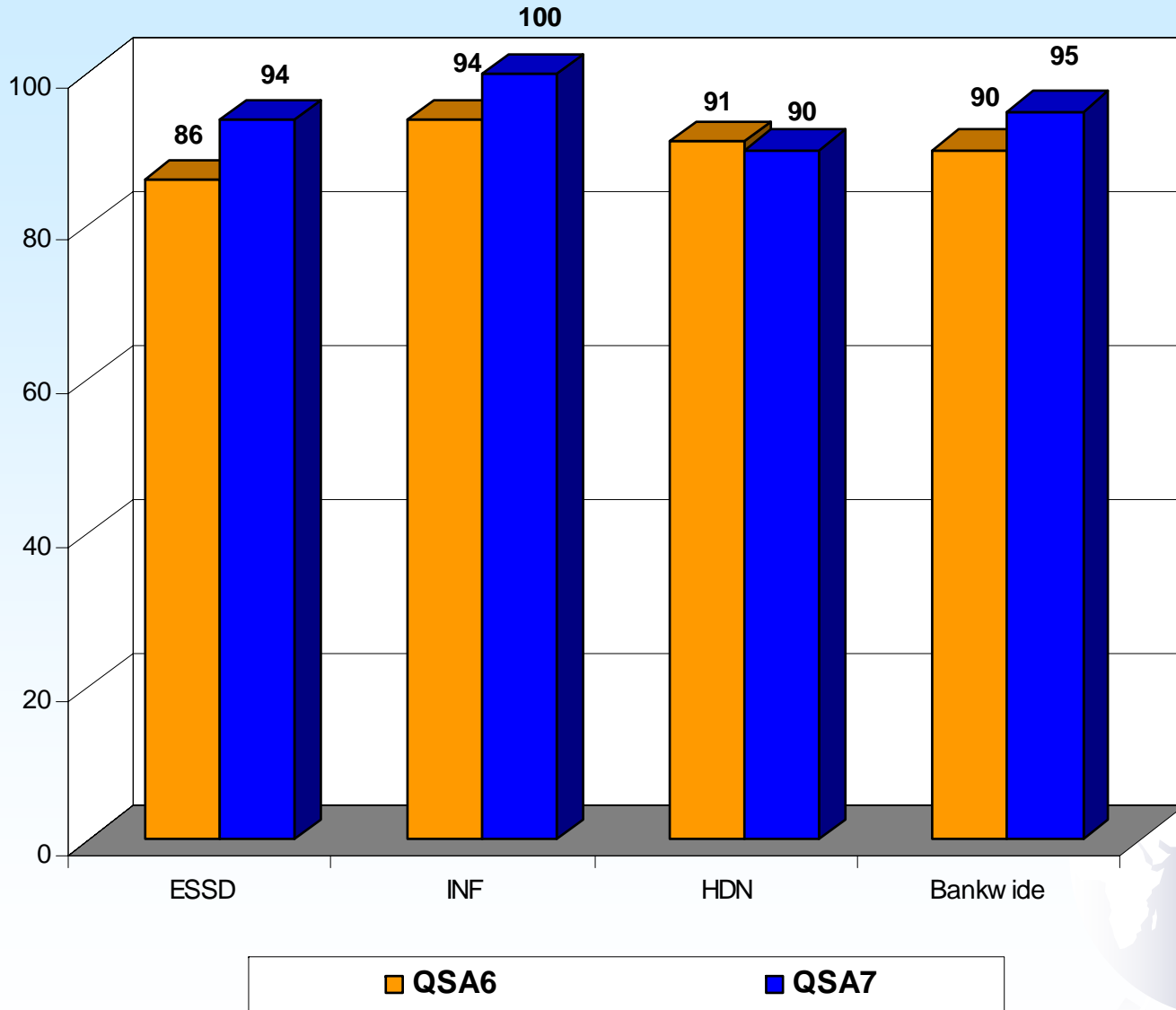


## Results by Network

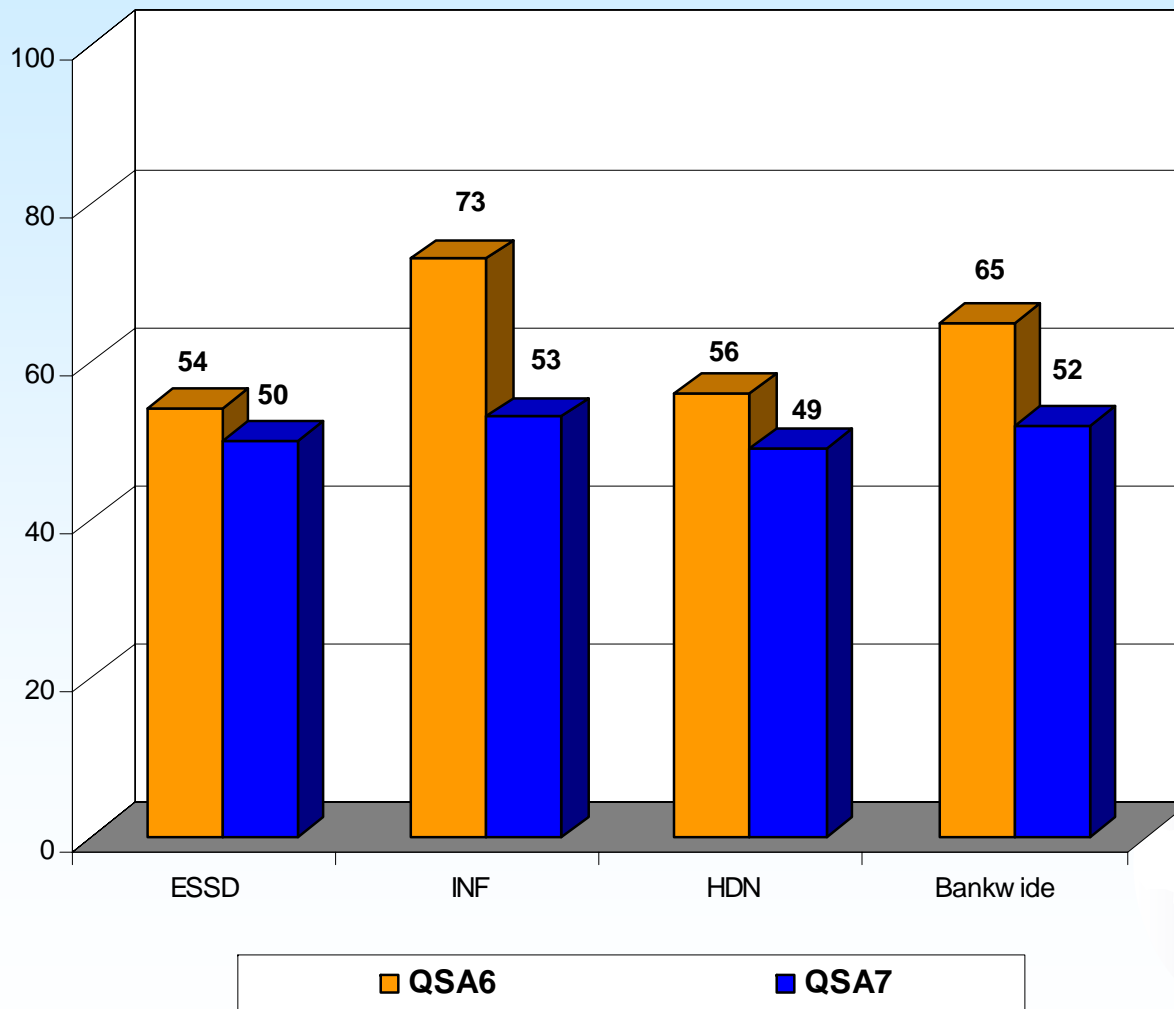
- Performance of networks with large cohorts in the QSA7 sample continues to be strong in terms of MS+ (Figure 9) with no significant differences in performance in terms of either MS+ or S+ (Figures 9 and 10). The main areas for improvement are:
  - In ESSD, Candor and Quality of ISR at 26% S+ is rated lower than other dimensions;
  - In INF, Focus on Development Effectiveness shows a decline from 73% S+ in QSA6 to 53% S+ in QSA7; and
  - In HDN, Supervision of Fiduciary/Safeguard Aspects at 42% S+ is below Bank-wide average of 60% S+.



**Figure 9: Results by Network**  
**(% Moderately Satisfactory or Better)**



**Figure 10: Results by Network  
(% Satisfactory or Better)**

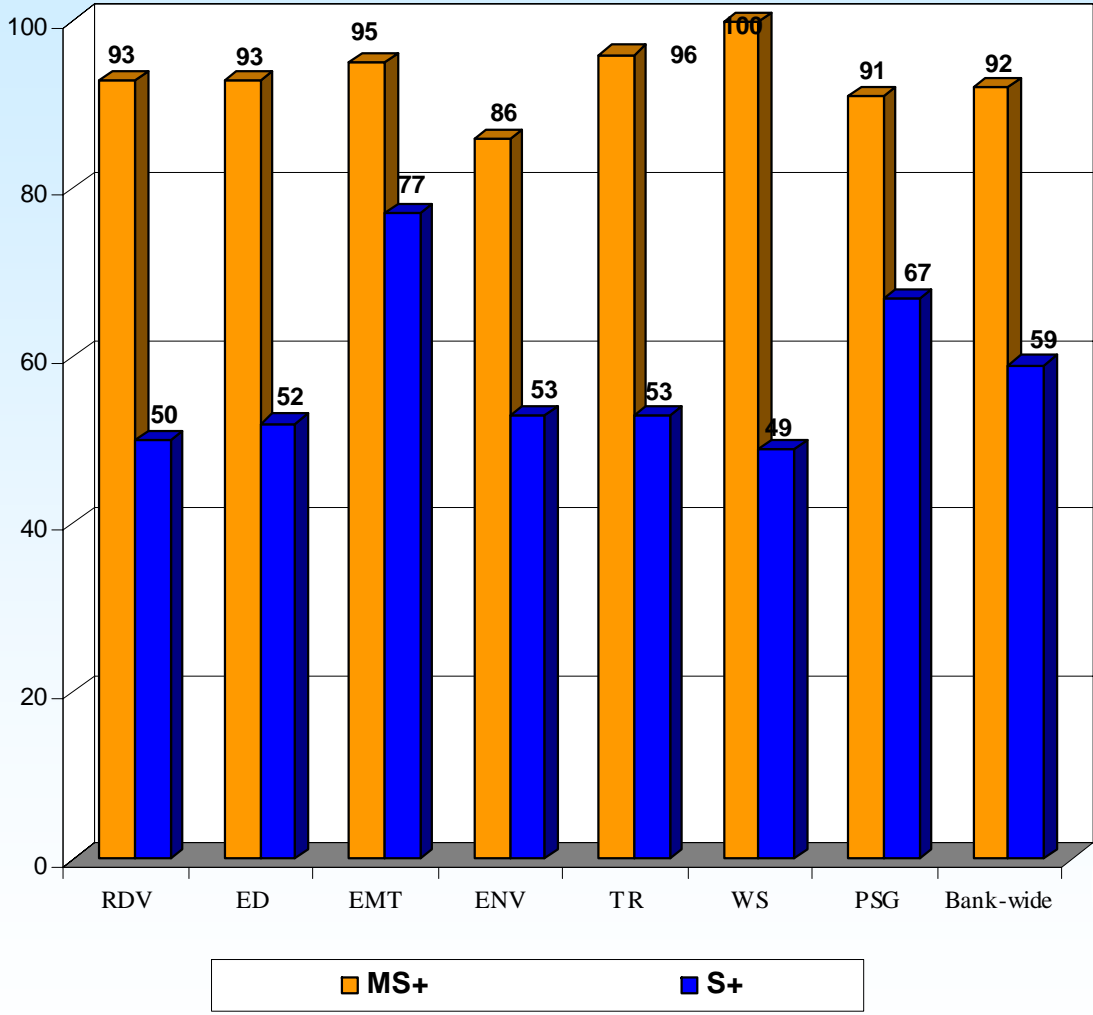


## Results by Sector Board

- The aggregation of QSA6 and QSA7 samples produced statistically robust results for seven sector boards (see Figure 11).
- There is convergence of overall performance among sectors in the MS+ category with an overall rating of 92% MS+ . However, among quality dimensions, Candor and Quality of ISR at 81% MS+ was the only dimension below 90% with Rural Development and Education at 75% MS+ and 76% MS+, respectively, the lowest among sectors for this dimension.
- There is a wide variation of performance among sectors in the S+ category, with Energy and Mining and Public Sector Governance outperforming other sectors.



*Figure 11: Results by Sector Board - Combined QSA6 and QSA7*



## Results by LICUS/non-LICUS

- The performance of LICUS projects at 88% MS+ was close to that of non-LICUS at 96%. In terms of MS+, performance of LICUS projects is comparable with non-LICUS in all quality dimensions except Candor and Quality of ISR where LICUS projects score lower.
- Notwithstanding the above, there is scope for improving performance of LICUS projects.
  - LICUS projects (42% S+) have lower scores than non-LICUS (65% S+) in Focus on Development Effectiveness due to weaker performance in Technical Assistance, Training and Capacity Building.
  - LICUS projects (48% S+) also score lower than non-LICUS (75% S+) in Adequacy of Supervision Inputs and Processes due to relative weaknesses in Staffing, Adequacy and Speed of Management Attention and Actions, and Adequacy of Supervision Budget.



# Results by Loan Size, Project Age and At Risk Category

## Loan Size

- There is no significant difference in quality of supervision by loan size.

## Project Age

- Projects in the 3-5 year age bracket – while at 89% MS+ - show decline from 65% to 41% in terms of S+. Attention to mid-year reviews and possible restructuring would help improve performance.
- Newer projects (less than three years old) show improvement in terms of MS+; this improvement is reflected in all dimensions.

## At Risk Category

- Actual problem projects show a decline in performance from 93% to 78% in terms of MS+ and 69% to 30% in terms of S+. This calls for greater focus on supervision of problem projects.



## IV. PORTFOLIO MANAGEMENT INDICATORS

- A cross-check of realism in supervision reporting provided by panelists' judgments reveals about twice as many problem projects in the sample as reported by task teams (Table 2).
- Panelists identify more than 3 times more risk flags on projects in the QSA7 sample than signaled by ISRs; M&E, project management, and financial management risk flags are especially underutilized (Table 3).\*

\* Panelists make a distinction between project performance and Bank supervision performance.



***Table 2: ISR vs. Panel Ratings of Problem Projects in QSA7 Sample***

	FY04		FY05		FY06	
	Region (%)	Panel (%)	Region (%)	Panel (%)	Region (%)	Panel (%)
<b>Problem Projects</b>						
APP (DO)	1	1	4	2	1	3
APP (IP/DO)	5	16	7	20	7	13
Sub Total APP (DO)	6	17	11	22	8	17
APP (IP)	4	8	4	10	4	13
<b>Total</b>	11	25	15	32	12	29



**Table 3: ISR vs. QAG Panel Risk Factor Ratings**

Risk Factor	FY04		FY05		FY06		FY04-06 *	
	PSR	QAG Panel	ISR	QAG Panel	ISR	QAG Panel	ISR	QAG Panel
	%	%	%	%	%	%	%	%
Legal Cov	3	13	10	18	9	16	8	16
Safeg	1	7	2	11	4	14	2	11
Financial Perf	14	19	NA	NA	NA	NA	14	19
M & E	5	34	8	38	6	39	6	37
Financial Mgmt	5	17	3	17	3	16	4	16
Proc	13	18	7	17	8	12	9	16
Project Mgmt	6	22	6	30	9	28	7	27
Critical Risks	7	13	NA	NA	NA	NA	7	13
Cntr Funds	NA	NA	6	15	5	12	6	13

\* Weighted Average using the number of projects rated on individual risk flags in FY04-06 fiscal years as the weights.



# V. RECOMMENDATIONS

## *Senior Management*

- ✓ Strengthen staff and managerial incentives and accountabilities for candor and quality of supervision reporting.

## *Regions*

- ✓ Improve quality and timeliness of Mid-term Reviews and encourage restructuring of problem projects as early as possible.
- ✓ Ensure that supervision budget allocations are tailored to project needs to maximize impact.
- ✓ Ensure that procurement and financial management staff are better integrated with supervision teams.

...Cont'd

## *Networks/Sector Boards*

- ✓ Provide guidance and training to task teams to improve the results framework of projects and promote dissemination of best practice examples across regions.

## *QAG*

- ✓ Explore options for improving the value added from future QSAs in light of progress made to date in supervision quality.



# VI. PREVIOUS QAG RECOMMENDATIONS

## Project-level Assessments

- One-third of operations in the QSA7 sample were previously assessed under QEA, QSA, or Country Lending. The extent to which Panels agreed with previous QAG assessments scored 74% in terms of S+ and 88% in terms of MS+ reflecting general agreement between current and past panels. Where there were disagreements, these were mainly in the following:
  - Previous overall assessment ratings should have been lower. In some cases, this is due to the 4 point scale used in the past;
  - Some previous Panels understated the problem of implementation readiness and did not identify implementation problems that could have been foreseen; and
  - Some previous Panels did not adequately identify weaknesses in the results framework in particular the KPIs.
- The extent to which task teams implemented previous QAG recommendations was rated 82% MS+ and 46% S+, which indicate scope for greater attention by teams to Panel recommendations. Implementing QAG recommendations would help improve the overall quality of supervision.



## Status of Follow-up on QSA6

RECOMMENDATIONS	STATUS	QAG Rating of Follow-up
<b>Senior Management</b>		
Address issues on candor and quality of project ratings and project restructuring.	<ul style="list-style-type: none"> <li>•Candor and quality of project ratings have not improved and remain the weakest quality dimension though some regions have begun addressing this problem;</li> <li>•Deterioration in Focus on Development Effectiveness continues to be a major concern in QSA7;</li> <li>• Recent Board approval of new restructuring procedures is expected to better enable project restructuring.</li> </ul>	Moderately Unsatisfactory
<b>Networks/Sector Boards</b>		
The ENV and GEF to develop remedial action.	There has been an improvement in the quality of supervision of GEF projects in QSA7.	Satisfactory
<b>Regions</b>		
SAR and MNA to prepare and implement action plans as needed.	SAR and MNA made significant gains in the MS+ category. However, both regions have much scope for improvement in the S+ category. Both regions have programs in place to further improve supervision quality.	SAR - Moderately Satisfactory MNA – Moderately Satisfactory

...Cont'd

## Status of Follow-up on QSA6

RECOMMENDATIONS	STATUS	QAG Rating of Follow-up
Establish/sharpen M&E for outcomes.	M&E for outcomes remains weak with a decline in timely identification and assessment of threats to DO. However, there has been progress in clarifying the process for revisiting outcomes and KPIs with teams now working on retrofitting these. QSA8 should review whether these initiatives are resulting in improved M&E and focus on development effectiveness.	Moderately Unsatisfactory
Rely on KPI for realistic performance rating and reporting of risks.	Performance rating and reporting of risks continue to be problematic, with Panels identifying two times as many problem projects and three times as many risk flags compared to ISR reporting.	Unsatisfactory
With ENV and SD sector boards, Regions to improve safeguards compliance.	There was slight improvement in compliance in SD but a worsening in ENV with more than double the number of non-compliance cases compared to QSA6.	SD – Satisfactory ENV – Unsatisfactory

...Cont'd

## Status of Follow-up on QSA6

RECOMMENDATIONS	STATUS	QAG Rating of Follow-up
<b>OPCS</b>		
Assist Regions standardize staff training on results framework including DO indicators.	There has been improvement in QSA7 from the situation in QSA6 on the basis of MS+. However, Quality and timeliness of data to support KPIs are rated 41% S+ indicating room for improvement.	Moderately Satisfactory
<b>QAG</b>		
Re-visit, in consultation with OPCS, the Regions, and IEG, the Portfolio Monitoring Indicators, to correct more drastically for consistent under-reporting of project risk and overstating pro-activity.	Done as part of FY06 ARPP. Revised definition for Realism Index is proposed to be effective by April 1, 2007.	Satisfactory
Tighten assessment methodology for Quality at Entry to give greater importance in arriving at the overall rating to results framework, realism of DOs, implementation readiness – noted as significant issues for about one-fourth of the projects in portfolio.	To be done as part of QEA8.	Not Rated
<b>OVERALL</b>		<b>Moderately Unsatisfactory</b>

## VII. FEEDBACK ON QSA7 PROCESS

- The TTLs, Panelists and Observers were requested to provide feedback on the QSA7 process. Response rates were: TTLs – 57%; Panelists – 63%; and Observers – 77%.
- The overall QSA7 process was rated 48% S+ and 80% MS+ by the TTLs, compared to 79% S+ and 88% MS+ in QSA6. Panelists and Observers rate QSA7 at above 90% in terms of both S+ and MS+.
- The strong points of the assessment were: panel skills mix, role of the moderator, time allotted for interview, and QAG management of logistics.
- The weak aspects of the assessment were: weaknesses in panel's familiarity with country background, low value added from panel interview, and lack of usefulness of the QSA as a learning experience.



- The TTLs rated at 92% or better in terms of MS+ the guidance and/or support from: Country Field Office, Sector Manager, and Procurement, FM, and Safeguards Units. Country Director/Manager (82% MS+) and Legal (86% MS+) received lower marks.
- 83% of the TTLs viewed the ISR as having value added for internal reporting but only 52% saw value added as an overall supervision tool. 65% were satisfied with the ease of use of the ISR.
- 28% of the TTLs reported internal or external pressures to upgrade project performance and risk ratings.
- Task teams reported spending 56 hours to prepare for and participate in the QSA7, compared to 66 hours for QSA6. Panelists reported spending 17 hours, with the feedback that more time was needed to perform the assessment.

