WORLD BANK OPERATIONS EVALUATION DEPARTMENT

Country Assistance Evaluation Retrospective

An OED Self-Evaluation

2005
The World Bank
Washington, D.C.

http://www.worldbank.org/oed

1818 H Street, NW
Washington, DC 20433
Telephone: 202-473-1000
Internet: www.worldbank.org
E-mail: feedback@worldbank.org

Manufactured in the United States of America
First edition May 2005

Cover photos courtesy of the World Bank Photo Library.

ISBN 0-8213-6316-6
e-ISBN 0-8213-6317-4

Library of Congress Cataloging-in-Publication data have been applied for.

World Bank InfoShop
E-mail: pic@worldbank.org
Telephone: 202-458-5454
Facsimile: 202-522-1500

Operations Evaluation Department
Knowledge Programs and Evaluation Capacity
Development Group (OEDKE)
Room # 19-352
MSN # 19-903
(202) 458-0382
Facsimile: (202) 522-3125

Printed on Recycled Paper

Contents

v Acknowledgments

vii Foreword

ix Executive Summary

xiii Acronyms and Abbreviations

1 **1 Introduction**
- 1 Background
- 2 CAE Retrospective Evaluation Approach

3 **2 Part I: What Have We Learned?**
- 3 Findings
- 5 Lessons Learned

11 **3 Part II: How Can the CAE Instrument Be Improved?**
- 11 Impact of CAEs on Country Strategy Formulation
- 13 Methodology
- 15 Ratings
- 17 Counterfactuals
- 18 Timing
- 19 Interaction with Country Teams
- 20 Interaction with Beneficiaries
- 21 Improving the Presentation of CAEs

23 **Annexes**
- 25 A: The CAE Methodology
- 29 B: Completed Country Assistance Evaluations, Fiscal Years 1995–2004
- 31 C: Summary of Board/CODE Interviews
- 35 D: Board CAS Discussions and CAS Documents
- 37 E: Summary of Bank Staff Interviews
- 41 F: Summary of Country Evaluation Methods Used by Other Agencies

43 **Attachment 1: Discussion Summary, CODE Subcommittee**

45 **Endnotes**

47 **References**

Box

17 3.1: Unsatisfactory Outcome Ratings Do Not Imply Poor Bank Performance

Figure

36 1: Distribution of Related References

Tables

4 2.1: Characteristics of the 25 CAE Countries

5 2.2: CAE Outcome Ratings and Project Outcome Ratings

5 2.3: Sectoral and Project Ratings for Fiscal 2001–03 CAEs

41 F.1: Use of Key Methodological Tools

Acknowledgments

This Country Assistance Evaluation (CAE) Retrospective was written by Poonam Gupta (task manager) and Kyle Peters (senior manager). Support and suggestions were received from René Vandendries, Chandra Pant, John Johnson, and Asita de Silva.

A number of external, independent evaluators and consultants were used to assist in the preparation of this self-assessment. Mike Hendricks evaluated CAE methods. Sharon Felzer prepared guidelines for the structured interviews in collaboration with Patrick Grasso (OED, Office of the Director). David Shirley interviewed CODE and Bank staff for this retrospective. Svenja Weber-Venghaus reviewed CAS transcripts, and Brandie Sasser and Weber-Venghaus surveyed the methods used by other donors in their country program evaluations. Janice Joshi and Agnes Santos provided administrative support.

Marcelo Selowsky of the Independent Evaluation Office of the International Monetary Fund and Ulrich Zachau (LCRVP) were peer-reviewers for this report. External peer review was provided by Vijay Kelkar and Faik Oztrak.

Acting Director-General, Operations Evaluation: *Ajay Chhibber*
Acting Director, Operations Evaluation Department: *R. Kyle Peters*
Senior Manager, Country Evaluation and Regional Relations: *R. Kyle Peters*
Task Manager: *Poonam Gupta*

Foreword

This report is a self-evaluation of OED's Country Assistance Evaluations (CAEs). In fiscal 2002, the Informal Subcommittee of the Committee on Development Effectiveness (CODE) discussed CAE methodology and the role CAEs play in assessing Bank interventions at the country level, as well as the need to continue to refine the methodology. This retrospective extends that discussion and addresses two questions: what have we learned, and how can the CAE instrument be improved? It addresses the first question by compiling lessons relevant for Country Assistance Strategies from the most recent CAEs, and the second by assessing the CAE process, methodology, and presentation.

This retrospective finds that an evaluation at the country level yields a more complete picture of the outcome of the Bank's assistance programs than individual project or program evaluations can offer. In a third of the country programs evaluated, aggregate outcomes from projects were satisfactory, but the country assistance program was unsuccessful overall. A review of recent CAEs suggests that the Bank needs to understand better the country context, deepen its country knowledge, and pursue adjustment or development policy lending only in situations where government ownership is strong. It also indicates that the outcomes of the Bank's country programs can be improved by undertaking more robust risk analysis, reducing planned assistance levels more sharply when faced with policy slippages, and lending more prudently in turnaround situations.

The second part of this retrospective is based on extensive interviews with CODE members, their alternates and advisers, and Bank Regional staff, as well as assessments of the CAE methodology and process by external evaluators. These assessments find that the methodological framework used in the past few years, while fundamentally sound, has some shortcomings. The structure tends to place too much emphasis on Bank instruments rather than on achievement of program objectives, and confusion has been created when different sections of the CAE appear to lead to different conclusions. These criticisms are consistent with the perceptions of CODE members who felt that CAEs focused too much on formal compliance issues and the impact of Bank instruments to the neglect of the Bank's development impact and the results achieved.

This retrospective concludes that the way forward is the adoption of a more results-based approach that clearly links the Bank's objectives and instruments with the outcomes achieved. This approach would outline the key objectives and results that the Bank has been trying to achieve in a country, based on the Country Assistance Strategies (CASs). For each objective and intended result, the CAE would evaluate the efficacy and relevance of the instruments used, as well as the relative contribution of the Bank (as opposed to that of others) to that objective. This approach would address the main recommendation given by the Bank's Board of Directors that the CAEs should focus more on the actual impact of programs at the country level.

The retrospective also examines several other issues that have arisen regarding CAEs.

First, it examines the debate surrounding CAE ratings. CODE members in general favored ratings, while the majority of Bank staff interviewed were critical of them. OED will review its use of ratings for the CAEs and is also discussing with management the standardization of self-ratings for Country Assistance Strategy Completion Reports (CASCRs) validated by an OED review. If such a system, akin to the project rating system, were introduced for all CASs, then ratings for CAEs might become redundant.

Second, the retrospective concludes that the timing of a CAE is crucial to its usefulness. OED will continue its policy of delivering CAEs to the Board and Bank staff about three to nine months ahead of planned CAS discussions.

Finally, this report notes a widespread perception that more interactions with country teams and with in-country beneficiaries during CAE preparation would enhance both the quality of the CAEs and the acceptance of CAE findings. OED, therefore, proposes to make an effort to ensure that in-country beneficiaries are consulted more uniformly and to involve country teams in midterm reviews of future CAEs to receive early feedback on emerging findings and recommendations.

Ajay Chhibber
Acting Director-General, Operations Evaluation

Executive Summary

This report is a self-evaluation of the Operations Evaluation Department's (OED) Country Assistance Evaluations (CAEs). In a fiscal 2002 discussion of OED's methodology for Country Assistance Evaluations, the Informal Subcommittee of the Committee on Development Effectiveness (CODE) noted the role CAEs play in assessing Bank interventions at the country level and the need to continue to refine the methodology. This retrospective responds to that discussion and has two primary purposes. First, it compiles lessons relevant for developing Country Assistance Strategies from the most recent batch of CAEs. Second, it assesses potential revisions to the CAE process, methodology, and presentation.

What Have We Learned?

The first part of this retrospective summarizes some major findings and lessons from the 25 CAEs completed during fiscal 2001–03, complemented by additional insights gained from CAEs completed in fiscal 2004. Overall, the lessons suggest that the Bank needs to understand the country context better, deepen its country knowledge, and better calibrate its assistance programs to observable progress on reform implementation.

This retrospective finds that an evaluation at the country level yields a more complete picture of the outcome of the Bank's assistance programs than do evaluations of individual programs or projects. In a third of the country programs evaluated, aggregate outcomes from projects were satisfactory, but the country assistance program was unsuccessful overall. This indicates the importance of assessing the country program in aggregate through a CAE. Otherwise, evaluation may not capture critical omissions in the country assistance programs; for example, that project interventions may not be addressing key development constraints.

The retrospective also notes that the Bank's success varied by sector. The outcomes in health and education have been relatively good. But in private sector development, public sector management, and rural development, the Bank's efforts have been less successful overall. Factors responsible for the unsuccessful outcomes in these sectors are:

- Reforms in these sectors seem to face stronger opposition from vested interests and potential losers.
- Institutional capacity constraints take time to resolve and many projects in these sectors attempt to improve upon the legal, institutional, and regulatory framework, but implementing change requires overcoming inertia and adverse incentives in the bureaucracies of many countries.
- These sectors suffer larger adverse effects from exogenous events and macroeconomic shocks

than others. Improving outcomes in all sectors would imply focusing more on measuring and supporting results-based indicators.

The retrospective has developed a number of lessons for improving the formulation and implementation of the Bank's Country Assistance Strategies.

First, successful country programs are tailored to the country context and an understanding of the political economy of reform is essential. Domestic politics and vested interests largely determine the pace and content of reforms in countries. The retrospective found a number of cases where an insufficient understanding of the political economy of reforms led the Bank to push reforms that stood little chance of success. More active dialogue with national governments, local government, and stakeholders enhances the Bank's understanding of political economy considerations.

Second, country knowledge is strongly associated with success. It is well understood that project interventions are more successful when they are based on in-depth analytical work. The same finding conveys to the success or failure of country programs. The retrospective finds that in more than two-thirds of successful country programs, the Bank's analytical work was timely. Analytical and advisory activities (AAA) can also be an effective vehicle for engaging governments in policy dialogue and informing civil society, but adequate attention needs to be paid to dissemination. In many cases, the attention paid to dissemination has been inadequate. This failing can also be a feature of participatory analytical work, as findings still may not be widely shared outside the government ministries that collaborated in the analytical work with the Bank.

Third, a number of CAEs provide clear evidence of the role technical assistance and investment loans can play in promoting institutional development and capacity building. But the sustainability of benefits requires that these operations be part of a broader macroeconomic stabilization and reform program. Linking technical assistance and investment loans with policy reforms supported by adjustment loans also improves the probability of success.

Fourth, adjustment lending can be successful, especially when combined with a strong government commitment to macroeconomic stabilization and structural reform; however, adjustment lending in the absence of sustained progress on stabilization and reform saddles the country with debt and weakens incentives to reform. The Bank needs to resist pressures to persist with adjustment lending in the absence of government commitment to reforms and a satisfactory track record in implementing them. The evidence from OED's evaluations of country assistance programs, however, provides only limited proof that the Bank resisted these pressures. In many countries, accommodating pressure to lend resulted in a number of unsatisfactory outcomes and, more important, weakened the incentive to reform.

Finally, the retrospective recommends several strategies that would improve the outcomes of Bank assistance programs:

- Undertake more robust risk analysis to carefully assess borrower commitment to reform and implementation capacity. This should be informed by the analysis undertaken by others and by the feedback obtained through wider dialogue with stakeholders.
- Reduce the level of planned assistance when faced with clear evidence of policy slippage. The Bank needs to set clear and meaningful triggers for its assistance programs and cut back its level of assistance when warranted.
- Lend more prudently in turnaround situations. This is especially the case in the presence of long-standing issues of implementation failure. Levels of assistance should be initially prudent and calibrated to measurable outcomes and meeting concrete benchmarks. The outcomes of Bank lending are better when lending goes up the ladder with the reform program, not ahead of it.

To summarize, optimistic projections or expectations with inadequate risk analysis often weaken the performance of country strategies. Programs should not be based on the best possible forecasted outcomes. Finally, country strategies need to be flexible, not rigid and narrow with only one path to follow.

How Can the CAE Instrument be Improved?

The primary purpose of an OED Country Assistance Evaluation is to draw lessons from the Bank's past activities to guide future strategies, illustrating the usefulness of the CAE instrument. Nevertheless, a Country Assistance Evaluation is a complex task and while many multilateral and bilateral agencies now undertake country program evaluations, there is no "standard" methodological approach. This second part of the CAE retrospective analyzes OED's current methodology and perceptions of the strengths and weaknesses of CAEs with a view to making further improvements in the methodology, evaluation process, and structure of CAEs.

OED formalized its methodology for CAEs in fiscal 2000 and the CAE approach became more standardized and consistent. The existing methodology starts with an evaluation of the relevance of the Bank's assistance strategy given the country's development needs and challenges, including whether any key development constraints were omitted. It then evaluates the assistance program across three dimensions: a bottom-up review of the Bank's products and services (lending, analytical and advisory activities, and partnerships); a top-down review of whether the Bank's program achieved its objectives or planned outcomes and had a substantive impact on the country's development; and analysis of the contributions of the Bank, borrower, other donors, and exogenous events to the outcomes.

For the retrospective, OED undertook a number of usage surveys. All objective measures—a tracer study of CAE recommendations, discussion of CAE findings and recommendations in subsequent Country Assistance Strategies (CASs), and Board discussions of subsequent CASs—indicated that CAEs are frequently referred to and their recommendations figure prominently in the Bank's CASs. Members of CODE, who were interviewed for this retrospective, felt that CAEs were useful, but were skeptical that Bank staff were using the CAE findings to the maximum extent. This was reflected in Bank staff views, which were on the whole much less positive, but were more positive when they perceived that a CAE had had an impact on the Bank's strategy. The conclusion reached in the retrospective is that the CAEs are playing a useful role, but revisions to the methodology, rating system, and interactions with involved parties would lead to a higher payoff in terms of acceptance and utilization.

OED's own assessment of CAE methodology and an independent external evaluator concluded that the methodological framework, while fundamentally sound, had several shortcomings. First, the structure tends to place too much emphasis on the Bank's instruments rather than on achievement of the program's objectives. Second, different sections of the CAE can appear to lead to different conclusions. For example, the outcome rating of the overall assistance program may deviate from the aggregate of project ratings, as noted earlier. Finally, the organization of the CAE tends to shift focus from the country to the Bank and back to the country, which makes it hard to follow the story line. OED's assessment is also consistent with the perceptions of CODE members who felt that CAEs focused too much on formal compliance issues and the impact of Bank instruments to the neglect of the Bank's development impact and the results achieved.

This retrospective concludes that the way forward is the adoption of a results-based approach, which clearly links the Bank's objectives, instruments, and the outcomes achieved. The approach would outline the key objectives and results that the Bank is trying to achieve in each country, based on the CASs. The strategy discussion would be based on a simple logic model that links inputs with expected results based on the information presented in the CAS itself and related documents. For each objective (and intended result), the CAE would evaluate the efficacy (and relevance) of the instruments used, as well as the relative contribution of the Bank (as opposed to other contributions) to that objective. This approach would address the Board members' main recommendation that the CAEs should focus on actual impact at the country level.

CAEs currently contain ratings of the outcome of Bank assistance, its sustainability, and institutional development impact. The CAE ratings have been the subject of considerable debate and criticism, which to some extent reflects the long-standing debate within the evaluation community on learning versus accountability. CODE members generally favored the ratings, arguing that they held the greatest potential for ensuring management attention to the CAE results. The majority of Bank staff were critical of the ratings, at times citing apparent inconsistencies between the ratings and the narrative, or, in other instances, citing a failure to rate different periods differently. But the key issue that arose is that many readers incorrectly perceive the outcome rating as either a rating of the country's development performance or as a rating of the Bank's performance. More needs to be done to make clear the nature and basis of the CAE ratings.

OED is reviewing the CAE ratings and is also discussing with management the regularization of self-ratings for CAS Completion Reports (CASCRs) validated by an OED review. If such a system, akin to the project rating system, were introduced for all CASs, then ratings for CAEs may become redundant.

The timing of the CAE is crucial to its usefulness. Bank staff prefer a CAE early in the CAS process, and CODE members prefer a CODE discussion close to the discussion of the relevant CAS. This retrospective found that the large majority of CAEs had been timed to appropriately inform CAS preparation and discussions. OED's review of the timing of the CAEs completed during fiscal 2001–03 indicates that more than 50 percent of the CAEs for which there have been subsequent CAS Board discussions had CODE discussions within six months of the CAS, and two-thirds had discussions within a year. Moreover, despite the perception of a lack of timeliness by Bank staff, all 22 of the CAEs discussed by CODE before Board discussion of the CAS were mentioned in those CASs. OED will continue its policy of delivering CAEs to the Board and Bank staff approximately three to nine months ahead of planned CAS discussions, but notes that the unpredictable scheduling of CAS discussions is a major challenge in meeting this performance target.

Finally, there was a widespread perception that more interactions with country teams during CAE preparation and with in-country beneficiaries would enhance both the quality of the CAEs and the acceptance of their findings. While this retrospective finds that there have been many more interactions with in-country beneficiaries than perceptions indicate, OED will make an effort to ensure more uniformly that in-country beneficiaries are consulted and their views incorporated. Similarly, the retrospective also finds that a more systematic interaction with country teams during the process of preparing the evaluation may help clarify and improve the acceptance of CAE findings and their incorporation in country programs. Therefore, OED proposes to involve the country team in midterm reviews of future CAEs to receive early feedback on emerging findings and recommendations.

Acronyms and Abbreviations

AAA	Analytical and advisory activities
AfDB	African Development Bank
APLs	Adjustable Program Loans
ARDE	Annual Review of Development Effectiveness
AROE	Annual Report on Operations Evaluation
CAE	Country Assistance Evaluation
CAS	Country Assistance Strategy
CASCR	Country Assistance Strategy Completion Report
CASPR	Country Assistance Strategy Progress Report
CDs	Country directors
CODE	Committee on Development Effectiveness
CPIA	Country Policy and Institutional Assessment
DAC	Development Assistance Committee
DFID	Department for International Development (U.K.)
EDs	Executive directors
ESW	Economic and sector work
FTB	Fast Track Brief
HIPC	Heavily indebted poor country
IBRD	International Bank for Reconstruction and Development
IDA	International Development Association
IDI	Institutional development impact
IDB	Islamic Development Bank
IMEP	Integrational Monitoring and Evaluation Plan
LICUS	Low-income countries under stress
LILs	Learning and Innovation Loans
NGO	Nongovernmental organization
OECD	Organisation for Economic Co-operation and Development
OED	Operations Evaluation Department
PSD	Private sector development
PSM	Public sector management
UNICEF	United Nations Children's Education Fund

Introduction

This report is a self-evaluation of the Operations Evaluation Department (OED) Country Assistance Evaluations (CAEs). In a 2002 discussion of CAE methodology, the Informal Subcommittee of the Committee on Development Effectiveness (CODE) noted the important role CAEs play in assessing Bank interventions at the country level and the need to continue to refine the methodology.

This retrospective responds to that discussion and has two primary purposes. First, it addresses the question "What have we learned?" by compiling lessons relevant for developing Country Assistance Strategies from the most recent batch of CAEs. Second, it assesses proposed revisions to CAE process, methodology, and presentation to answer the question "How can the CAE instrument be improved?"

Background

The focus of the Bank's development assistance is now firmly anchored at the country level. Country Assistance Strategies (CASs) have become the centerpiece of the Bank's country programs and place the Bank's operational work in the context of the country's own strategy and development objectives. The Bank is currently strengthening its strategy formulation process through the adoption of a results framework for CASs to define more clearly the country results that the Bank seeks to influence, as well as linkages between these results and the program supported by Bank initiatives.

In keeping with the Bank's shift from a project to a country focus, OED first began evaluating country programs in fiscal year 1995. The methodological approach evolved, building upon OED's project methodology. By fiscal 2000, the methodological approach was formalized,[1] and in fiscal 2002, as noted above, the CAE methodology was discussed at a CODE subcommittee meeting. As a result, the CAE approach became more standardized and consistent.

The existing methodology starts with an evaluation of the relevance of the Bank's assistance strategy given the country's development needs and challenges, examining, among other matters, whether any key development constraints were overlooked. It then evaluates the assistance program across three di-

Country Assistance Strategies (CASs) have become the centerpiece of the Bank's country programs.

mensions. The first is a bottom-up review of the Bank's products and services (lending, analytical and advisory activities [AAA], and aid coordination), which are used to achieve Bank objectives. The next step is a top-down review of whether the Bank's program achieved its objectives or planned outcomes and had a substantive impact on the country's development. The first and second steps test the consistency of findings regarding the product and services and development impact dimensions. Finally, the CAE attributes responsibility for the country strategy outcomes to the Bank, the borrower, other donors, and exogenous events.

By the end of fiscal 2004, OED had issued CAEs assessing the impact of Bank assistance to 64 countries.[2] These evaluations cover roughly 47 percent of borrowers and 76 percent of total gross commitments. This retrospective focuses primarily on the 25 CAEs completed in fiscal 2001–03, but also includes some insights gained from the additional eight CAEs completed during fiscal 2004.

Five joint CAEs have been undertaken with other multilateral institutions to foster the sharing of experiences and evaluation harmonization. The Lesotho and Rwanda CAEs were prepared jointly with the African Development Bank (AfDB); the Jordan and Tunisia CAEs were prepared jointly with the Islamic Development Bank (IDB), and the Peru CAE was prepared in parallel with the Inter-American Development Bank.

The primary purposes of these joint exercises have been to promote learning on both sides and encourage the development of common evaluation standards, consistent with the broader harmonization agenda of the Bank and other donors. For example, the AfDB recently revised its guidelines for Country Assistance Evaluations, closely aligning them with OED's existing methodology. Joint country evaluations have also been effective in identifying key constraints and gaps in donor assistance. For example, the AfDB evaluation of its Rwanda program highlighted the lack of a post-conflict policy. Finally, joint country evaluations are one avenue to reduce the burden of multiple, separate donor evaluation efforts on recipient countries. But these benefits are partially offset by transaction costs and added time needed for consultation and coordination.

CAE Retrospective Evaluation Approach

To achieve the two objectives of the retrospective, OED commissioned a number of background studies to assess the lessons learned, the CAE methodology, and the usage of CAEs by the Board, CODE (and the CODE subcommittee), and Bank staff.

- OED staff reviewed the 25 CAEs completed between fiscal 2001 and 2003 to ascertain lessons applicable across CAEs. The analysis and findings of this review are included in Part I.
- OED commissioned an independent consultant to conduct qualitative research, consisting of in-depth interviews with Board members, CODE and CODE subcommittee members, and other participants in CAE discussions at CODE or subcommittee meetings. In total, 30 people were interviewed.
- Two tracer studies were undertaken. Transcripts of Board discussions and written statements by executive directors about CASs or CAS Progress Reports (CASPRs) were reviewed to assess the impact of CAEs on subsequent CAS discussions. Also, desk reviews of 25 CASs (or CASPRs) produced following the CAEs were analyzed to determine the timeliness of CAEs and the utilization of their recommendations.
- OED commissioned the same independent consultant who conducted the CODE interviews to conduct a similar qualitative exercise with Bank staff. In-depth interviews were conducted with 20 current and former country directors associated with one or more of the 25 CAEs completed in fiscal 2001–03, five current and former Regional vice presidents, and 18 additional Bank staff.
- A professional evaluator examined the current CAE methodological approach and proposed methodological changes, which would, in particular, focus CAEs on results related to CASs.
- A desk review was undertaken of country evaluation methods of other bilateral/multilateral agencies to identify good practices.

Part I: What Have We Learned?

This chapter summarizes the findings and lessons from the 25 CAEs completed in fiscal years 2001–2003. It also includes some insights gained from the additional 8 CAEs completed during fiscal 2004.[3]

These 25 CAEs are not a random or a representative sample. CAE countries are selected based on several factors. The most important is the timing of the Bank's next Country Assistance Strategy (CAS), but other factors are also considered, such as OED's ability to collaborate with other international financial institutions, Regional balance, and whether a previous CAE had already covered a country. So, while the countries cannot be called a random sample of Bank borrowers, these 25 CAEs constitute a reasonable basis for this "findings and lessons" discussion (see table 2.1).

There is a good mix of low- and middle-income countries, of International Bank for Reconstruction and Development (IBRD) and International Development Association (IDA) borrowers, and of Regions. Four post-conflict countries are included. The sample also contains a good mix of countries with outcome ratings that were satisfactory (about three-fifths), and of those with ratings that were unsatisfactory. It is fairly evenly divided between good performers and weak performers, as measured by the Bank's country policy and institutional assessment (CPIA) rating. As noted in the *2003 Annual Review of Development Effectiveness* (ARDE), there is a positive correlation between CAE outcome ratings and policy performance (OED 2004a, page 15). One weakness is that the coverage of Latin American and South and East Asian countries is scant, primarily because evaluations of countries in these Regions were fairly complete prior to fiscal 2001.

Findings

CAE outcome ratings can deviate from the aggregate of project outcomes. In about a third of the CAEs, the outcome rating of the Country Assistance Strategy was unsatisfactory, but the aggregation of project outcomes in the country during the CAE period was satisfactory (table 2.2). This is not surprising. The CAE is a comprehensive evaluation of the Bank's program in a country that comprises both projects and analytical and advisory activities. Moreover, CAEs must make an assessment of overall Bank strategy, including size, sectoral composition, and type of lending. For example, the CAE outcome may be unsatisfactory if there are critical omissions in the Bank's overall assistance strategy, even if the outcomes of individual projects are rated satisfactory.

CAE outcome ratings can deviate from the aggregate of project outcomes.

CAEs found that Bank programs were more successful in education and health

Table 2.1 Characteristics of the 25 CAE Countries

Income level	
Low	10
Lower-middle	12
Upper-middle	3
	25
IBRD, IDA, or Blend	
IBRD	14
IDA	8
Blend	2
	24[a]
Post-conflict or not	
Yes	4
No	21
	25
CPIA[b] level	
Above average	13
Below average	11
	24[c]
Region	
Africa	12
East Asia and the Pacific	2
South Asia	1
Europe and Central Asia	5
Middle East and North Africa	3
Latin America and the Caribbean	2
	25
OED outcome ratings	
Highly satisfactory	2
Satisfactory	9
Moderately satisfactory	8
Moderately unsatisfactory	3
Unsatisfactory	9
	31[d]

a. For West Bank and Gaza, the financing comes through a World Bank Trust Fund and is on IDA or grant terms.

b. CPIA = Country Policy and Institutional Assessment.

c. No CPIA for West Bank and Gaza.

d. Four countries received different ratings for subperiods.

than in other sectors, and least successful in private sector development, rural development, environment, and social protection (table 2.3). These results are also largely reflected in aggregate project ratings for the same set of 25 CAEs (table 2.3). The only significant divergences between sectoral ratings in CAEs and project outcome ratings were in the areas of private sector development (PSD), rural development, and public sector management (PSM), where the relatively unfavorable outcomes in the CAEs are in contrast to the high proportion of satisfactory outcomes at the project level.

The divergence in success ratings at the project and sector levels in the areas of private sector development, rural development, and public sector management could reflect several factors. First, the political economy of reforms in these sectors is more problematic, and opposition from potential losers and vested interests is likely to be more focused and sustained. Individual projects may be successfully implemented, but it will take a longer and more sustained effort to overcome political opposition and achieve successful outcomes at the macro level. For example, in transition economies, the historical antipathy to private business on the part of the *nomenklatura* has strong ideological roots and is not easily overcome. Despite legal and regulatory reforms, there remains an ingrained bias against private business in many of these countries. In the less-developed economies of Asia, Africa, and Latin America, major rural development initiatives often encounter conflicts of interest among powerful groups (such as big farmers versus laborers, or rural dwellers versus urban population) that take time and political acumen to resolve.

***Second,* even if there was no strong political opposition, these areas are crucially dependent on institutional reforms and capacity building, which take time and are not always captured even in the time frame of a CAE.** For example, private sector development depends on an effective judicial and court system, which takes time to develop. Civil service reform takes time to implement and the results take even longer to materialize. Thus successful outcomes at the sector level will take time to materialize.

***Finally,* outcomes in these areas are more dependent on economy-wide developments**

and exogenous factors. The growth of the private sector depends not just on the legal and regulatory framework for private sector development, but also on progress in other areas of economic policy such as macroeconomic stabilization, infrastructure, and the financial sector. Noneconomic conditions within the country, such as law and order, and external events, such as prevailing market sentiment in the region, also play a major role in private sector development. Rural development and reforms in the public sector are also influenced significantly by economywide developments. (This is not to say that outcomes in other sectors do not depend to some extent on such factors. It is a matter of degree.)

Table 2.2 CAE Outcome Ratings and Project Outcome Ratings

	CAE outcome ratings (%)	
Country portfolio outcomes	**Satisfactory**	**Unsatisfactory**
Satisfactory	53	33
Unsatisfactory	7	7

Source: World Bank database.

Lessons Learned

Lesson one: An understanding of the political economy of reforms, including government commitment and ownership of reforms and the degree of political support for or opposition to them, is essential to develop realistic Country Assistance Strategies, specific assistance programs and projects, and analysis of risks.

Economic reforms often failed either because the government was not committed to them, or because the government underestimated opposition to reforms and was unable to carry them through. An insufficient understanding of the political economy of reforms and the nature of the state may have led the Bank, in some cases, to push reforms that stood little chance of success. For example, in Zimbabwe, one of the principal motivations behind the Government's economic and social policy was to ensure indigenous ownership of productive assets. A proper appre-

An insufficient understanding of the political economy of reforms and the nature of the state may have led the Bank, in some cases, to push reforms that stood little chance of success.

Table 2.3 Sectoral and Project Ratings for Fiscal 2001–03 CAEs

Sector	Project outcomes (% satisfactory)[a]	CAE sector outcome ratings (%)[b]			
		Satisfactory	Unsatisfactory	Mixed	Not assessed
Education	100.0	48	8	4	40
Health	86.5	40	20	8	32
Social protection	37.9	12	24	20	44
Environment	25.5	16	16	24	44
Rural	87.2	12	36	28	24
Financial sector	81.6	24	16	32	28
Infrastructure	87.7	36	16	24	24
Public sector	99.9	24	32	28	16
Private sector development	60.1	12	24	44	20

a. This is the percentage of satisfactory project outcomes by sector (commitment value) for the countries where CAEs were completed during fiscal 2001–03.

b. These columns are ratings by CAE task managers and a desk review of CAEs of the outcomes by sectors. In most cases, explicit ratings by sectors were not included in the CAEs and thus, these ratings should be treated as indicative only.

Source: World Bank database and OED staff estimates.

Institutional reforms and capacity building for effective governance are critical to successful outcomes.

ciation for the importance of this issue might have led the Bank to give priority to land reforms. The Government's reluctance to undertake parastatal and civil service reform would also have been better understood. Privatization and civil service reforms in Peru were threatened by opposition from the middle class. Their opposition could have been lessened by a different sequencing of reforms and social protection measures to mitigate the adverse consequences of privatization and civil service reforms.

Lesson two: Institutional reforms and capacity building for effective governance are critical to successful outcomes.Since capacity building takes time, these reforms need to start early, and be followed through over several years.

Institutional development is at the core of development effectiveness. Stronger institutions are associated with a 20-percentage point increase in the likelihood of a project's outcome being rated satisfactory (OED 1999). Successful reform outcomes are often undermined by weak institutions or the absence of important institutional arrangements.

For example, in the transition economies of central and eastern Europe and the former Soviet Union, the existing institutions were not designed for a market economy. As the command economy collapsed, and market-oriented reforms were implemented, there was a lag before the right institutions developed, which prevented the full benefits of the reforms from emerging.[4] Institutional reform goes beyond changing organizational structures and rules; it also involves the discarding of long-established habits and patterns of behavior—a complicated and lengthy process. Major institutional changes may have to be spread over several years.[5]

ESW must be timely, of good quality, and be fully integrated into the design of Bank strategy, programs, and projects.

Lesson three: Successful outcomes from reforms in a sector often depend on complementary reforms and success in other sectors. The sequencing and packaging of reforms need to take account of this complementarity.

Several CAEs noted that more successful outcomes could have been achieved had more attention been given to the complementarity of reforms in different sectors. Private sector development cannot be assured simply through privatization. It also depends on reform of the public sector through deregulation, changes in taxation policy and administration, and anti-corruption reforms. For example, in Mongolia, important measures were implemented to promote private sector development—such as enabling private property, removal of price and margin controls, reduction in trade barriers, simplification of tax regime, and improved framework for foreign investment—but not enough was done to lighten the heavy hand of the state on the economy. Inefficient government regulatory and oversight functions continued to hamper private sector development.

As several CAEs noted, the growth of the private sector depends very much on an effective judiciary. The effectiveness of financial sector reforms in improving financial intermediation has also been shown to depend critically on reforms in the enterprise sector and the public sector. For example, experience shows that as long as enterprises are being bailed out by the public sector (through explicit or implicit subsidies and build-up of tax arrears), they have less incentive to borrow from banks. And on the supply side, banks are reluctant to lend to enterprises that survive only because they are subsidized by the public sector. Either way, financial intermediation is stunted.

Lesson four: ESW must be timely, of good quality, and be fully integrated into the design of Bank strategy, programs, and projects. Timely ESW is particularly important for first-time or renewed borrowers and for stop-go reformers.

ESW can play an important role in developing the Bank's assistance strategy and in enhancing its effectiveness. Two-thirds of the CAEs that reported favorable outcomes also reported that the ESW was timely. The same CAEs reported unfavorable outcomes when the ESW was not.

In Brazil, timely ESW helped the Bank to direct lending toward education, health, and rural poverty projects in the Northeast with maximum impact on poverty reduction. And in Vietnam, ESW demonstrated that rural poverty was strongly associated with inadequate economic infrastructure, helping orient the Bank's lending program toward economic infrastructure.

Examples of less satisfactory contributions of ESW include the case of Kazakhstan, where a social protection project sought to mitigate the social impact of privatization by strengthening the institutional capacity of unemployment services to streamline procedures for registration and payment of unemployment benefits. The poverty assessment that came two years later showed that policies to facilitate labor mobility and to equip workers for changed circumstances were required, rather than a strengthening of capacity of unemployment services. In Bulgaria, a timely poverty assessment might have enabled a social protection loan to more effectively address targeting of social assistance to the needy.

In a few countries, ESW was timely and of high quality, but findings were either not fully taken into consideration in designing strategies, or used selectively, thus reducing their relevance and effectiveness. For example, in Jordan, Bank lending during the 1990s did not address the high level of government expenditures, despite analytical work identifying how these expenditures could be reduced and better targeted. In Peru, the Financial Sector Reform Loan (fiscal 1999) did not address fundamental problems in the sector, which had been correctly identified in an ongoing assessment of the financial sector.

ESW can build a good knowledge base and help the Bank engage governments in its policy dialogue. It can also assist a government to understand the costs and benefits of reforms and help design its reform program. The long-term impact of ESW should also be borne in mind, as it initiates debate on policy issues and serves to familiarize civil society with reform issues. But it goes without saying that ESW cannot catalyze policy change when governments lack commitment to reform.

Lesson five: Specific investment and technical assistance loans can be useful vehicles for promoting institutional reforms, but for benefits to be sustainable, these operations should be part of a broader stabilization and economic reform strategy, and must be supported by strong governmental commitment to reform.

About 22 of the 25 CAEs provide strong evidence of the potential for specific investment lending and technical assistance loans in promoting institutional development in both low-income and middle-income countries.[6] Twenty of these countries also received adjustment loans. In many countries, specific technical assistance and institution building loans were linked closely to adjustment loans, helping build capacity to formulate and implement policy changes supported in adjustment loans.

Strong government commitment and clearly articulated priorities remained a key factor in the effectiveness of these instruments in promoting institutional development. In this environment the longer time frame of investment loans allowed the Bank to build relationships with counterparts, and to combine advice with financial assistance especially to sector ministries. Conversely, when strong and sustained government commitment to institutional reform is absent, Bank assistance is not likely to be successful.[7]

Lesson six: Adjustment lending in the absence of sustained progress on reforms only saddles the country with additional debt and can weaken incentives for future reforms.

Adjustment lending can be successful, especially when combined with a strong government commitment to macroeconomic stabilization and structural reform. However, the Bank should resist pressures to persist with adjustment lending in the absence of the government's commitment to reforms, and a satisfactory track record in implementing them. The rationale for adjustment loans was to provide financing to alleviate the cost of adjustment that occurred when structural reforms were implemented.

Adjustment lending in the absence of sustained progress on reforms only saddles the country with additional debt and can weaken incentives for future reforms.

However, the review of CAEs for this retrospective showed that adjustment lending was appropriately delayed in only two countries (Bulgaria and Brazil) when reforms stalled. In as many as seven other countries, the Bank went ahead with adjustment loans even though little progress was being made towards the Bank's assistance objectives and corporate goals.[8]

In these countries there were pressures to lend for a variety of reasons: to exploit a unique "window of opportunity" (Kenya), show support to the government (Morocco), maintain relationships (Zimbabwe), avert a potential crisis (Peru), avoid a return to communism (Russia), and prevent negative net flows (Zambia). While these factors may well have influenced the Bank, there is little doubt that in some of these countries the pressure to reform may have been further diluted by the Bank's decision to lend, saddling these countries with debt.[9]

Lesson seven: Thorough, hard-headed, and realistic risk analysis is important to increase the realism of country strategies.

The Bank needs to be more realistic and hard-headed in its country assessments and country strategy formulations, including assessing a borrower's commitment to reforms and its implementation capacity, receptiveness to Bank advice, and the impact of reforms on growth and poverty alleviation. Consistent errors of over-optimism on the part of the Bank regarding borrower receptiveness to Bank advice, willingness to undertake difficult reforms, and government's capacity to implement reforms has plagued Bank strategies in many of the countries evaluated (Haiti, Jordan, Kazakhstan, Kenya, Kyrgyz Republic, Lesotho, Morocco, Paraguay, Peru, Zambia, Zimbabwe). This optimism often persisted in the face of contrary evidence[10] and contributed to lending decisions by the Bank that failed to meet their objectives.[11]

Country assistance strategies in many coun-tries assumed a much stronger growth performance than warranted by past country experience or experience of other countries facing similar constraints and prospects. Evaluations pointed to unrealistic growth projections in several countries (Jordan, Kazakhstan, Kyrgyz Republic, and Zambia). Unrealistic growth estimates contributed to inappropriate Bank assistance strategies and entailed real costs for the countries. Had growth estimates been realistic, the Bank would have likely concentrated more analytical work on the sources and constraints to growth, as well as on poverty reduction measures. Realistic growth projections would also have illustrated more clearly debt sustainability issues and greater efforts would have been made to seek debt relief or other forms of concessional financial assistance, in order not to impose too high a burden of external debt;[12] and countries may have been persuaded to undertake deeper reforms to accelerate economic growth.

The Bank should be more prepared to reduce the level of planned assistance when faced with clear evidence of policy slippages.

Lesson eight: The Bank should be more prepared to reduce the level of planned assistance when faced with clear evidence of policy slippages.

In addition to identifying risks, the Bank needs to be prepared to modify its assistance program to reflect wavering government commitment or policy slippage. A positive example is Bulgaria in the mid-1990s, where lending was scaled down and a major adjustment loan was put on hold in the face of rising macroeconomic risks and a lack of government commitment to address reform issues. But the review of CAEs suggests this did not happen in a number of cases. In Peru, the Bank's program did not contain triggers to reduce lending if risks materialized. In Kenya, the 1998 strategy also recognized risks and the program contained CAS benchmarks, but the bank did not follow through when the benchmarks were not met. In Morocco, the Bank provided a policy reform support loan in the late 1990s as a way of rewarding the country's movement toward a more open political system and commitment to reform. The loan was too unfocused to have a major impact on any of the critical reform areas identified in the country strategy. Many of the actions taken prior to Board presentation were first steps, sometimes in the form of studies or plans, and many others did not show concrete results.

Lesson nine: The Bank should be especially cautious in turnaround situations. Realistic country assessments rather than wishful thinking should inform its assistance strategy.

Typically, the Bank has difficulty in identifying "turning points" and in calibrating its response to changing country conditions. It tends to react slowly to deterioration and too quickly to improvements. In the Dominican Republic, for example, the Bank failed to recognize an upturn during fiscal 1992–95 and failed to support the government during a crucial period of successful economic reform. This mistake cost the Bank its influence in the country. The Bank failed to recognize the downturn in Peru in 1997 and continued the support it was rendering earlier when reforms were implemented. In Kenya, initial steps toward reforms in mid-2000 were prematurely identified as an upturn and rewarded with increased lending; reforms ultimately stalled.

To overcome this, the Bank needs to keep its ears closer to the ground through its Resident Missions, its contacts with civil society, and relevant ESW work to gain a better understanding of the political economy. As far as possible, a government's commitment should be assessed on the basis of its track record in implementing reforms, not declarations of intent. Levels of assistance should be initially prudent and calibrated to measurable outcomes and meeting concrete benchmarks. This is especially the case in situations where there are long-standing issues of implementation failures.

Typically, the Bank has difficulty in identifying "turning points" and in calibrating its response to changing country conditions. It tends to react slowly to deterioration and too quickly to improvements.

Part II: How Can the CAE Instrument be Improved?

The primary purpose of an OED CAE is to draw lessons from the Bank's past activities to guide future strategies. Part I of the retrospective synthesized a number of lessons and findings generated by the most recent set of OED CAEs. These lessons, which accord well with the findings of other OED evaluations, illustrate the usefulness of the CAE instrument. Nevertheless, a CAE is a complex task, and while many multilateral and bilateral agencies now undertake country program evaluations, there is no "standard" methodological approach.[13]

This second part of the CAE retrospective analyzes OED's current methodology and perceptions of the strengths and weaknesses of CAEs with a view to improve the methodology, evaluation process, and structure of CAEs. This chapter discusses the impact of CAEs on country strategy formulation, the rating system, the use of counterfactuals, the timing of CAEs, interactions with country teams and in-country beneficiaries, and the presentation of CAEs. Each section describes the perception of CAE users, the results of tracer studies and OED's own review, and concludes with suggestions for improvements.

The analysis in this chapter is based on survey research regarding the usage of and perceptions of CAEs by the Bank's Board of Directors and Bank staff, as well as the assessment of an external evaluator, who reviewed the CAE methodology and a set of recently completed CAEs. As noted in the introduction, a number of surveys and tracer studies have been conducted to inform the analysis. Detailed findings from the surveys of CODE members are contained in Annex C; from the analysis of CAS Board discussions in Annex D; from the interviews of Bank staff with familiarity of CAEs in Annex E; and from the analysis of the methodology of country program evaluations by other agencies in Annex F. Each annex details the objectives of the survey and describes the participant sample, as well as the views and findings.

Impact of CAEs on Country Strategy Formulation

Earlier studies have indicated that CAEs do influence Bank Country Assistance Strategies and there is consistency between CAE recommendations and CASs. OED's 2002 Annual Report on Operations Evaluation (AROE)[14] undertook a survey of Bank country teams on CAEs. About 71 percent of respondents thought

Recent analysis confirms that CAE recommendations are well reflected in subsequent CASs.

that the CAEs had influenced Bank policies at the country level, 80 percent thought that a CAE is relevant to their work, and 93 percent said that if they were to start work on a new country, they would find a CAE to be quite useful. For the same AROE, OED carried out a tracer study to assess the extent to which CAE recommendations on a country were incorporated into the subsequent CAS, and were reflected in the assistance actually delivered. The study reviewed 15 CAEs, all completed in fiscal 1999 or earlier to ensure that enough time had passed to permit a review of assistance actually delivered. The study found strong consistency between CAE recommendations and the CAS, reinforcing the message from the survey.

Recent analysis confirms that CAE recommendations are well reflected in subsequent CASs. For this retrospective, OED undertook a new tracer study for the 25 CAEs completed during fiscal 2001–03. This analysis reviewed the CASs which immediately followed the CAEs to assess to what extent the formal recommendations included in the CAE's management action records are supported. Of the 22 CAEs for which subsequent strategy documents have been prepared,[15] two-thirds of the CAE recommendations are fully supported in the Country Assistance Strategies. Another quarter are partially supported and fewer than 10 percent were not supported at all. Also, all of the 22 CASs contained some reference to the relevant CAE; in 16, the CASs contained extended discussions of CAE findings and recommendations. Thus, it does appear that the CAEs are having an impact on the Bank's strategy formulation, although it is not possible to attribute the support for any individual recommendation to the analysis contained in the CAE.

CAEs also appear to have had an impact on Board discussions of subsequent CASs, and the issues raised in CAEs are aligned with Board discussions.

CAEs also appear to have had an impact on Board discussions of subsequent CASs, and the issues raised in CAEs are aligned with Board discussions. For this retrospective, OED reviewed Board transcripts from 13 CAS discussions, for which the most recent CAEs were available as of April 2003.[16] About one-third of chairmen's statements, either written or oral, explicitly referred to the CAE at least once. In total there were 85 references to CAEs in the 13 discussions. About a quarter (21) of these were very general references, while the remaining 75 percent (64) were issue-oriented. References to CAEs were more frequent during full CAS discussions or when the outcome rating was unsatisfactory; chairs representing Part I constituencies made the largest number of references to CAEs. In total, references to CAEs comprised only a small fraction (6 percent) of the multitude of issues raised during CAS Board discussions. Nevertheless, even when a CAE was not referenced, there was alignment of the issues raised in the CAEs and the Board discussions of the respective CASs and Progress Reports.

There is a divergence between the results of the tracer surveys and the perceptions by the Board and Bank staff of CAE usage and usefulness. The usage surveys indicate that CAEs are known to both Board members and Bank staff. CODE members interviewed for the retrospective noted that they used CAEs to prepare for CAS Board discussions, to place sectoral and thematic issues in a country context, and to assess CASs. Some CODE members cited specific instances in the interviews when CAEs positively influenced Bank country strategies. Nevertheless, a majority doubted that CAEs exerted a significant influence on management decisions or policy issues, noting that a lack of endorsement by senior management of CAE findings served to undermine their impact. This "lack of endorsement" was apparent in interviews with Bank staff, including regional vice presidents and country directors. Their perceptions of the usefulness of CAEs were less positive, but were more positive when they perceived that a CAE had had an impact on CAS preparation. Bank staff identified several factors inhibiting the effectiveness of CAEs: a lack of timeliness, inadequate consultation with country teams, and a lack of useful insights or lessons for Bank involvement in a country.

Reducing the divergence and improving the utilization of CAEs could be enhanced by changes to the methodology, presenta-

tion and timing of CAEs, as well as to the frequency of interactions with country teams. Despite the fact that CAEs figure prominently in CASs and CAS Board discussions, our surveys raise a number of issues about CAEs. Board members expressed some skepticism over whether operational staff were utilizing CAE findings to the maximum extent possible. Bank staff have also raised concerns about the underlying methodology, the timing of CAEs and the balance between accountability and learning in their content. All these factors require careful review, and, where appropriate, resolution, so that CAEs can have a higher payoff in terms of utilization and acceptance of their findings and recommendations. The remainder of this Chapter will address the most salient issues identified from the usage surveys: the CAE methodology; CAE ratings; use of a counterfactual; timing; interaction with country teams; interaction with in-country beneficiaries; and CAE presentation.

Methodology

The adoption of a formal methodology for CAEs in fiscal 2000 has led to a more consistent approach to Country Assistance Evaluations.[17] The CAE methodology uses the Country Assistance Strategy (CAS) for its initial evaluative framework, supplemented by judgments about whether the strategy omitted addressing any key development constraints. The CAE gauges the extent to which the CAS's major strategic objectives were relevant and achieved, without any shortcomings.

The current CAE methodological approach is to examine the country assistance program across three dimensions:

- A *Products and Services Dimension,* involving a "bottom-up" analysis of major program inputs—loans, AAA, and partnerships/aid coordination
- A *Development Impact Dimension,* involving a "top-down" analysis of the principal program objectives for relevance, efficacy, outcome, sustainability, and institutional impact
- An *Attribution Dimension,* in which the evaluator assigns responsibility for the program outcome to four categories of actors: (a) the client; (b) the Bank; (c) partners and other stakeholders; and (d) exogenous forces (e.g., events of nature, international economic shocks, and the like.).

Reducing the divergence and improving the utilization of CAEs could be enhanced by changes to the methodology, presentation and timing of CAEs, as well as to the frequency of interactions with country teams.

With the introduction of this formal methodology, CAEs became more similar in structure and content. Each CAE begins with an examination of the overall development situation and the relevance of the CAS objectives. This is followed by an in-depth evaluation of all products and services used to achieve Bank objectives. The third section evaluates whether the Bank's objectives have been achieved and contains the CAE ratings. The contributions of the different partners to the results obtained are assessed, and the CAEs conclude with recommendations and lessons.

Despite the adoption and dissemination of a formal CAE methodology, interviews revealed a perception that the CAE methodology could be improved and made more transparent and consistent. A majority of the CODE participants felt that CAEs were consistently of high quality and contained substantive analysis of the Bank's country programs. Nevertheless, concerns were raised about inconsistencies in the findings and variations in quality. CODE members expressed concerns that the documents are sometimes too focused on monitoring compliance with formal institutional objectives to the neglect of the Bank's development impact on the country. They felt that a CAE should focus on the results achieved and on how instruments were employed to achieve the results. Bank staff also raised concerns about the CAE methodology. They felt that the methodology was not clear, standardized, consistent, or streamlined. Moreover, they contended that too few CAEs were generating findings and lessons that provided pragmatic guidance for future program management. Some Bank staff expressed a preference

The CAE structure tends to place more emphasis on the Bank's instruments (lending, AAA, aid coordination) than on the impact or results of the Bank's interventions.

for evaluations focusing on a limited number of specific sectors, rather than attempting to evaluate an entire country program.

OED's own assessment also finds that the current CAE methodology could be modified in several ways to increase transparency, consistency, and readability. To determine the accuracy of the perceptions collected from CODE and Bank staff, OED contracted an independent, external evaluator to undertake an evaluation of the current CAE methodology with no restrictions or preconditions. Program evaluation methods adopted by other multilateral and bilateral donors were also reviewed.[18] From this analysis and the usage surveys, a number of observations emerged. The current three-dimensional approach to CAEs has created a consistent format for reporting, so that over time the reader knows what to expect in each chapter. But this methodological approach also appears to have several shortcomings:

- *The structure tends to place more emphasis on the Bank's instruments (lending, AAA, aid coordination) than on the impact or results of the Bank's interventions.* CAEs give a greater emphasis to program inputs than to the program's development impact. Of the 18 CAEs with clearly similar structures, 13 devoted more than two and a half times as much attention to inputs than to the development impact.
- *Different sections of the CAEs can appear to lead to different conclusions.* As noted in Part I, in one-third of the CAEs, satisfactory average portfolio ratings did not correlate with a satisfactory outcome rating for the program as a whole. While the explanation is fairly straightforward, (that is, an individual project or ESW evaluation tends to focus on the project's stated major objectives without necessarily determining how relevant those objectives were to achieving the overall program goals in force at that time) the CAE structure is not conducive to assessing the underlying reasons for this disconnect.
- *The structure shifts focus from one section to the next, which makes it hard to follow the story line.* The current CAE structure alternates from a focus on the client (first section), the Bank (second section), and back to the client. It is, therefore, difficult to follow the link between objectives (which are discussed in the first section) and the outcomes (which are discussed later in the third section), and it is difficult to see the linkage between the Bank's inputs and services (which are evaluated in the second section) and specific outcomes.
- *CAE terminology lacks clarity and consistency.* OED's internal review of CAEs found that the terminology used in different CAEs at times is seemingly inconsistent, as different task managers may interpret the same concept somewhat differently. For example, the distinction between Bank performance and the outcome of the Bank's program is unclear in some CAEs, which can be a consequence of Bank strategies focusing on higher-level objectives and failing to clearly articulate a set of intermediate objectives. Another example is the treatment of attribution, which is inherently a complex concept in a country program.

The way forward is the adoption of a results-based approach, which clearly links the Bank's objectives, instruments, and the outcomes achieved.

The way forward is the adoption of a results-based approach, which clearly links the Bank's objectives, instruments, and the outcomes achieved. Results should be the organizing structure of the CAE and the evaluation of the Bank's instruments (products and services) should be seen simply as a means to test whether the end results, which are the Bank's key objectives, have been achieved. This is consistent with the results-based framework that is being adopted in Bank strategy formulation and is also consistent with the methodology being adopted by other donor agencies. The approach would be to outline the key objectives and results that the Bank is trying to achieve in each country, based on the CASs. The strategy discussion would be based on a simple logic model that links inputs with expected results based on the information presented in the CAS itself and related documents. For each objective (and intended result), the CAE would evaluate the efficacy (and relevance) of the instruments used, as well as the relative contribution of the Bank (as opposed to other contributions) to that objective. This approach would be results-based and would address the Board members' main recommendation that the CAEs should focus on actual impact at the country level.

A number of recent CAEs have begun to adopt aspects of this new methodology. For example, the Zambia CAE identified four main themes (objectives) by which the Bank planned to reduce poverty: (i) promoting a stable macroeconomic environment; (ii) privatizing state-owned enterprises; (iii) targeting assistance directly to the poor and vulnerable groups, and (iv) improving public sector efficiency and governance. The CAE then reviewed the Bank's activities and evaluated the results in each of the four areas separately, which it then summarized into its overall ratings. The Brazil CAE identified and evaluated progress on the key objectives of poverty alleviation and sustainable growth, and also identified and evaluated sub-objectives under each objective. More recently in fiscal 2004, the China, Tunisia, Bosnia-Herzegovina, and Croatia CAEs all adopted a more objective-based evaluation approach.

OED's review of CAEs and of the methodological approach of other donors suggests that changes are also needed in the attribution concept in the CAEs. The Development Assistance Committee (DAC) definition of attribution involves establishing a causal link between a change and a specific intervention. This is a difficult task partly because of the inherent limitations of proving causality in a naturally occurring setting. Countries are not laboratories and CAEs are not reporting on experiments; controlling for other influences and isolating influences is extremely difficult in the real world. The concept of "most likely association" captures much better the reality of the link between the Bank or a donor's actions and observed outcomes than does "attribution." This would imply establishing to the degree possible how and to what extent the Bank's interventions were related to the results achieved. The Brazil CAE and the Bank's focus on the Northeast is a good example of the concept of most likely association as a way to assess the Bank's contribution.

CAEs rate the outcome of Bank assistance, its sustainability, and institutional development impact.

OED will intensify its efforts to ensure a more consistent application of the methodology. OED already has a rigorous internal review process that helps ensure consistency across evaluations. Nevertheless, to improve consistency, OED will develop a glossary of terms, similar to the "2002 Glossary of Key Terms in Evaluation and Results Based Management" produced by the Working Party on Aid Evaluation of the Development Assistance Committee of the Organisation for Economic Co-operation and Development (OECD). Finally, for all large-client CAEs, outside reviewers will continue to be used. These external reviewers, who are distinguished retired Bank experts, outside academics, and client country development policy officials, help to bring in outside views on the Bank's programs and to ensure the validity of CAE findings.

Ratings

CAEs rate the outcome of Bank assistance, its sustainability, and institutional development impact. Currently, OED country evaluations contain explicit ratings for country programs; OED is the only multilateral or bilateral agency that rates country programs in this way. There are three ratings in most CAEs.[19] First, CAEs rate the outcome of the Bank's assistance program. The outcome rating gauges the extent to which major strategic objectives were relevant and achieved, without any shortcomings. Outcome is rated on a six-point scale, ranging from highly satisfactory to highly unsatisfactory. Second, most CAEs rate the institutional development impact (IDI) of the assistance program. The four-point rating of IDI (high, substantial, modest, or negligible) measures the extent to which the program bolstered the client's ability to make more efficient, equitable, and sustainable use of its human, financial, and natural resources. Finally, OED rates sustainability or the resilience to risk of the development benefits of the country assistance program. Sustainability is rated as highly likely, likely, unlikely, highly unlikely, or, if available information is insufficient, nonevaluable.

CAE ratings have been the subject of much debate and criticism. Underlying this criticism is a long-standing debate within the evaluation community over which of the two complementary goals should be made uppermost: establishing accountability for the results, or promoting learning about how to do things better in the future.

This historic debate comes out in the Board interviews, where the majority of participants regarded ratings as indispensable, with some arguing that ratings offer the greatest potential for holding management accountable for its actions and ensuring that it pays appropriate attention to the findings. However, some did feel that less attention should be paid to the ratings, arguing that disputes over ratings create defensiveness among operational staff and actually impede learning from the program experience. Nevertheless, most Board interviewees concurred that the rating scales needed to be clarified, with individual ratings definitions made more transparent and mutually exclusive.

The majority of Bank staff interviewed were also critical of the rating system, at times citing inconsistencies between the ratings and the narrative used to justify them, and a failure in some instances to rate different implementation periods individually. Staff also argue that a "one-size-fits-all" rating standard fails to take sufficient account of crucial differences in client capabilities, initial endowments, and exposure to exogenous risks. Views were divided as to whether ratings should be ignored, de-emphasized, or improved with more rigorous criteria.

A key issue is that many readers incorrectly perceive the outcome rating as either a rating of the country's development progress or as a rating of Bank performance. CAEs rate the outcomes of Bank assistance programs, but it is clear from the interviews of Board participants and Bank staff, as well as written comments on CAEs, that the nature of the outcome rating is not clearly understood. An assistance program is evaluated on how well it met its particular objectives, which are typically a subset of the client's development objectives. Situations can arise where the outcome of the Bank's assistance program is fully satisfactory yet the country's economy is deteriorating, or vice versa. Therefore, there is a distinction between the outcome rating and the client's overall development outcomes. Also, as the outcome rating evaluates how well the assistance program met its objectives, it is not synonymous with Bank performance, which CAEs do not rate. As clearly set out in the CAE methodological framework, the outcome of the Bank's assistance program is determined not only by the Bank's performance but by the joint impact of four agents: (a) the client; (b) the Bank; (c) partners and other stakeholders; and (d) exogenous forces (e.g., events of nature, international economic shocks, etc.). This helps explain the seemingly contradictory finding in some CAEs that the Bank's performance may have been good but the outcome unsatisfactory. For instance, in Bulgaria (1991–97), Haiti (1997–2000), Paraguay (1989–99) and Rwanda (1990–93) outcomes of the Bank's assistance strategy were rated unsatisfactory but the CAEs note that the Bank performed well (see Box 3.1).

Many readers incorrectly perceive the outcome rating as either a rating of the country's development progress or as a rating of Bank performance.

OED will make changes to help clarify the ratings and make them more transparent and consistent. OED intends to continue to rate outcome and sustainability. The revised methodological approach outlined in this retrospective will greatly facilitate an unbundling of the outcome rating. In the future, the CAEs will disaggregate the outcome rating, providing separate and distinct sub-ratings of the results achieved for each key objective. These sub-ratings will then be aggregated to arrive at a summary outcome rating using weights implied by the strategy. Modifications are also proposed for the ratings of sustainability. The concept may be recast as the "likelihood of sustainability," transforming it into a rating of whether the ongoing benefits of the country program are subject to high or low risks. Since most Bank activities now have institutional development as an integral part of their objectives, OED intends to drop the separate IDI

Box 3.1 Unsatisfactory Outcome Ratings Do Not Imply Poor Bank Performance

Bulgaria (1989–97): **The objectives of the Bank's strategy of supporting macroeconomic stabilization and growth, private sector development, and poverty alleviation were not met but the Bank appropriately focused on stabilization and debt reduction. When reforms stalled, the Bank put on hold a major adjustment loan until a more appropriate environment existed for reforms and supported institutional changes and specific sectoral needs through investment lending.**

Haiti (1986–00): **Bank assistance had an unsatisfactory impact on challenges facing Haiti (poor governance, inadequate public sector capacity, lack of donor coordination, poor educational quality and access, and badly deteriorated infrastructure). Until 1997, assistance by the Bank and other donors overwhelmed the government's absorptive capacity. They resumed traditional types of assistance programs, without first overcoming the enormous governance and institutional barriers to economic recovery and program implementation. Internal Bank staffing issues also reduced the effectiveness of the program in Haiti. Since 1997, Bank performance has improved. Irregularities in the road project were cleaned up. In a situation where it could not lend and where borrower performance continued to be problematic in the ongoing program, the Bank cancelled the remaining balances. It is, appropriately, maintaining a watching brief and undertaking reviews of key sectors.**

Paraguay (1989–99): **Bank assistance objectives of private sector development, agricultural development/natural resource management, and poverty alleviation/human resource development were not met. The Bank prepared well-focused ESW. Its investment in dissemination stimulated public debate, particularly on the need for reform. The 1997 CAS focused on short-term objectives that were feasible within Paraguay's institutional and policy constraints. It recognized that the risks identified in 1993 were still largely in place and that the deteriorating political situation made reform unlikely. It recommended a cautious funding strategy with a maximum of one or two projects approved per year subject to improved implementation of the existing portfolio.**

Rwanda (1990–93): **The objectives of the Country Assistance Strategy of growth with equity, creation of an enabling environment for private sector development, protection of the environment, and improvements in public resource management capacity were not met. There were some shortcomings in Bank performance (e.g., project design did not adequately reflect human and financial constraints confronting the government) but the Bank performed well overall. It understood the fundamental economic difficulties facing Rwanda and tried to persuade the Government, albeit unsuccessfully, to change its long-standing policies and give a greater role to the private sector, reduce public subsidies to failing enterprises, and improve the quality of social services.**

Source: **Bulgaria, Haiti, Paraguay, and Rwanda CAEs.**

rating and capture it as part of the outcome rating. Finally, as experience is gained with results-based CASs and CCRs, which will also contain ratings, OED will review once again the rating system in CAEs and make further modifications as needed. If the Bank adopts a system of assessing CAS performance through ratings in CAS Completion Reports—akin to the ratings for individual projects—OED may drop the ratings of CAEs and turn them more into a lesson learning review.

Counterfactuals

CODE and Bank staff views diverge on the usefulness of counterfactuals. Less than a third of participants in Board interviews credited counterfactuals with generating useful CODE subcommittee discussions. More than half of the CODE respondents criticized the counterfactuals for frequently presenting a too rosy or utopian picture of the Bank's role in the country, complaining that CAEs tend to exaggerate the role of the Bank and suggest that if the Bank had not been involved, the country would have done much worse. In contrast to CODE members,

More than half of the CODE respondents criticized the counterfactuals for frequently presenting a too rosy or utopian picture of the Bank's role in the country.

Bank staff in CAE discussions, interviews and memos criticized CAEs for not having an effective counterfactual showing what alternative strategy the Bank could have followed, or what would have happened in the absence of the Bank; or, alternatively, for not taking into account the often marginal role of the Bank in the country's development. A review of Bank staff memos, however, showed that these views were expressed only when a CAE was critical of the Bank's assistance. The counterfactual issue was not raised where the CAE noted that the Bank performed well.

OED recognizes the importance of counterfactuals for illustrating what the Bank could have done differently in specific situations. OED recognizes the need for a counterfactual but it is useful to emphasize that this task has been difficult partly because of the inherent difficulties of proving causality in real world situations. That is why only 10 of the 25 CAEs reviewed (40 percent) attempt a discussion of the counterfactual, and these generally focus more on what the Bank could have done differently than on what would have happened in the absence of the Bank. The Russia CAE, for example, notes that it is difficult to construct plausible counterfactuals in the case of such a complex program, where much of what the Bank did was in response to explicit shareholders' expectations. Instead the CAE describes how the Bank tackled difficult challenges and, with the benefit of hindsight, identifies cases where it may have been more appropriate to adopt a different stance.

The proper role of a counterfactual analysis in the CAE is to compare actual outcomes with the ex ante analysis contained in the Bank' Country Assistance Strategy. The CAE methodological approach judges a CAS by what it intended to achieve. Therefore, a counterfactual already exists against which the CAEs can assess what actually happened. CASs typically provide multiple scenarios, matched by a shifting set of planned Bank instruments. So, OED has a ready counterfactual, which avoids the complexities of trying to guess what would have happened in the absence of the assistance program. In other words, the CAS itself often provides the counterfactual. The quality of such embedded counterfactuals is expected to improve significantly now that the Bank has set uniform guidelines for producing results-based CASs. Comparing what was expected ex ante in the country strategies with what actually happened and whether the divergences were reasonable is a useful way of conceptualizing the counterfactual. Where the CAS does not provide a fully defined counterfactual, other, less attractive options may be utilized, such as comparing: (i) the client country's development progress during the implementation period of the CAS to that of other countries; (ii) the development progress in specific sectors or geographic areas, which logically should have been significantly impacted by the program interventions; and (iii) intermediate outcomes logically affected by program interventions.

Timing of the CAE is essential to its effectiveness, but there is some tension between Bank staff and the Board about when a CAE is perceived as useful.

Timing

Timing of the CAE is essential to its effectiveness, but there is some tension between Bank staff and the Board about when a CAE is perceived as useful. There was widespread agreement among all users of CAEs that the timing of a CAE was an essential factor in its usefulness. A CAE that is completed either too far in advance of a CAS or after the completion of a CAS is viewed as having a limited impact. This is confirmed by OED analysis of the usage of CAEs during Board discussions. In effect, the impact of a CAE is diminished significantly if it comes either too early or too late in the CAS review process. For example, if the time elapsed between CAE distribution and Board discussion of the CAS is a year or more, the CAE seems to have little influence on the deliberations. On the other hand, a majority of CODE interviewees also cited cases in which CAEs were distributed too close to the date of Board CAS review to have any meaningful impact. Operational staff perceive that the delivery of CAEs is not sufficiently predictable to be a useful strategy development tool. Even Bank staff who received their CAEs in a timely fashion expressed skepticism about their future usefulness, because of their awareness of other instances in which other CAEs were made available too late in the process to inform the CAS. CODE interviewees stated they preferred receiving the CAE one to three months before a CAS discussion. Bank staff, on the other hand, indicated that, in order to inform the CAS, at least an early draft of the CAE would have to be available by the upstream review stage, which in a number of cases would be more than six months prior to a Board discussion.

A large majority of recent CAEs appear to have been timed appropriately to inform CAS preparation and CAS discussions. A key criterion in OED's decision to undertake a CAE in a particular country is the CAS cycle. Other factors are considered—regional balance, collaboration with other IFIs, and the long lead times required for CAEs in major Bank borrowers—but the CAS cycle is the primary determinant of the decision to schedule a CAE. OED's review of the timing of the 25 CAEs completed during fiscal 2001–03 indicates that, on the whole, CAEs have been timely, but there is scope for improvement. Of the 25 CAEs, eight were completed within three months of the CAS discussions, four others were completed between four and six months, and another four were completed in the year prior to the CAS discussion. In three cases, no subsequent CAS has yet been prepared and only in six cases were the CAEs completed more than one year prior to the CAS discussion. In these cases, CAS schedules were changed. Furthermore, despite the perception of Bank staff that CAEs were not timely for CAS preparation, all of the CAEs discussed by CODE before Board discussion of the CAS were mentioned in those CASs. Of the 22 CAEs for which CAS discussions have been held, CAE findings and recommendations were explicitly incorporated in 16 CASs, and the other six CASs either included a box referring to the CAE (3) or mentioned the existence of the CAE (3).

OED will endeavor to time CAEs to support CAS preparations and discussions. OED will continue its long-standing policy of delivering CAEs to the Board and Bank staff around three to nine months ahead of planned CAS discussions, whenever feasible. One very important contribution CAE users could make would be to increase the predictability of the Board schedule of upcoming CAS discussions, as some of the timing issues recounted above resulted from long delays due to rescheduling CAS discussions. CODE can assist in this process by ensuring that the 'green sheet' of the CAE discussion is finalized and made available to the Board prior to discussion of the corresponding CAS. In any event, whether the CAS is delivered on time or not, OED believes CODE would benefit from scheduling discussion of the underlying CAEs without waiting for the CAS. OED will also work to increase interactions with country teams during the CAE preparation process to ensure that evolving CAE findings are discussed with country teams during the early stages of CAS preparation (this is discussed in the next section). This should enable OED to meet both the needs of operational staff and the Board in terms of timing CAE deliveries.

Interaction with Country Teams

Bank staff voiced a clear preference for more interaction with OED staff during the preparation of the CAE. Bank staff, particularly country directors, who provided a positive assessment of CAEs, stressed the value and importance of the process through which the CAE was generated. Participants who assessed the CAEs positively were more likely to describe the CAE—and its value for their work—in terms of an ongoing learning process and repeatedly described open and regular communications between the country team and the CAE mission team. Conversely, Bank staff who assessed the CAE negatively described a low level of interaction between the country team and the CAE mission team. CODE interviewees also noted that the CAE should not only be seen as a stand-alone document, but also as a tool for facilitating interactions between Bank staff and OED in order to improve the receptivity of CAE findings and recommendations with Bank management.

More interaction with country teams is a vehicle to improve receptivity to CAE findings. OED's review of the CAE process did reveal that too many CAEs were prepared with little interaction with country teams. In some instances, the country director was asked to comment on the approach paper, was inter-

More interaction with country teams is a vehicle to improve receptivity to CAE findings.

viewed by the CAE task team and then, only months later, was presented with a draft CAE. While individual members of the country teams were always interviewed individually for each CAE, there have been few instances of systematic interactions with the country team and country director during the course of the evaluation. OED proposes to involve the country team in midterm reviews of future CAEs to receive early feedback on emerging findings and recommendations. This approach would create a vehicle to assure country teams that their views are valuable and are being heard; to demonstrate sensitivity to the country team's need for understanding and managing the potential impact of a negative evaluation on ongoing operations, and to increase the receptivity of country teams to the evaluative findings.

Interaction with Beneficiaries

There is a widespread perception that CAEs rely too heavily on interviews with Bank staff and government officials and desk reviews of documents, and place too little emphasis on the views of in-country beneficiaries. Many CODE interviewees expressed concerns about the excessive reliance on secondary research conducted exclusively in the Washington office rather than working directly with government officials and other in-country beneficiaries. While most CODE interviewees specifically acknowledged the inherent difficulties of involv-ing government officials and other country beneficiaries in the evaluation process—limited budget and resources at OED, lack of institutional capacity at the country level, and the potential to confuse or disrupt existing Bank relationships—the majority insisted that OED should continue to strive for greater country involvement in the CAEs. Bank staff also felt that CAE task teams relied excessively on secondary resources, had inadequate consultations with country beneficiaries, spent too little time in-country, and did most of the work as a desk review. They perceive that important people, who were in key positions at the time when policies were being made, were not adequately consulted.

OED will make efforts to ensure that CAEs reflect more fully the views of civil society on the Bank's country programs.

OED's review does indicate a reliance on the views of Bank staff and government officials, but the evidence shows that the majority of CAEs go well beyond desk reviews. A review of CAEs that listed people interviewed indicates that the majority of people interviewed by CAE teams are either past and present government officials, from the World Bank Group, or other donors (see Annex D). Fewer consultations were held with the private sector, NGOs, country academics, and other client stakeholders. One explanation for this is that project performance assessment reports constitute a key building block of CAEs and they are prepared after extensive consultation with project beneficiaries, thereby lessening the need for CAE teams to consult with them. The evidence also clearly shows that CAEs do not rely solely on secondary research and desk reviews. They incorporate the views of government officials, both past and present, in the country, and do not omit the views of key policy makers from the past who were involved in influential decisions.

OED will make efforts to ensure that CAEs reflect more fully the views of civil society on the Bank's country programs. OED will work toward involving in-country beneficiaries more in the evaluative process. One approach that was successful in the Brazil CAE was the use of in-country experts and academics as part of the CAE team. This need might also be addressed through wider use of qualitative methods, such as small focus groups and informal workshops with governmental officials and other stakeholders. However, such methods are frequently costly, and thus, will be used in a manner commensurate with OED's budget.

Improving the Presentation of CAEs

CODE interviewees feel that changes in the document design of the CAEs would improve their readability. More than half of the CODE interviewees stated that they had difficulty reading and finding relevant information in CAEs. For the most part, this is less a problem of thoroughness and length than of organization. One common suggestion was that the section on "Findings and Recommendations," which is most likely to be read by management and the Board, be strengthened and moved to the front of the report.

Reviews of Board discussions reveal that certain parts of the CAE are more heavily referred to than others. OED's examination of the issue-oriented CAE references in Board discussions suggests very selective reading of the CAEs (see Annex D). One third of the references point to the "Recommendations" section of a CAE, and another half referred to the CAE's "Memorandum to the Executive Directors and the President" or to the issues raised at the CODE or Subcommittee meetings. Only a few references come from the main text. This underlines the importance of the summary-type sections of the CAE and of ensuring that discussions of the CAE with the CODE and Subcommittee is effective and available to the full Board for the CAS discussion.

OED proposes to experiment with the design of the document to enhance readability. One possibility is to give greater prominence to the findings and recommendations in the document's overall organization and design. Since most Board members read primarily the summaries, cross-referencing would assist readers in finding desired information (which they may otherwise wrongly assume is missing from the document) and in encouraging readers to explore and familiarize themselves with portions of the document that they may not read thoroughly.

ANNEXES

ANNEX A: THE CAE METHODOLOGY

This methodological annex describes the key elements of OED's Country Assistance Evaluation (CAE) methodology.[20]

CAEs evaluate the outcomes of Bank assistance programs, not clients' overall development progress, An assistance program needs to be assessed on how well it met its particular objectives, which are typically a subset of the client's development objectives. If an assistance program is large in relation to the client's total development effort, the program outcome will be similar to the client's overall development progress. However, most Bank assistance programs provide only a fraction of the total resources devoted to a client's development by donors, stakeholders, and the government itself. In CAEs, OED evaluates only the outcome of the Bank's program, not the client's overall development outcome, although the latter is clearly relevant for judging the program's outcome.

The experience gained in CAEs confirms that program outcomes sometimes diverge significantly from the client's overall development progress. CAEs have identified assistance programs that had:

- Satisfactory outcomes matched by good client development
- Unsatisfactory outcomes where clients achieved good overall development results, notwithstanding the weak Bank program
- Satisfactory outcomes where clients did not achieve satisfactory overall results during the period of program implementation.

Assessments of assistance program outcome and Bank performance are not the same. An unsatisfactory assistance program outcome does not always mean that Bank performance was also unsatisfactory, and vice-versa. This becomes clearer once we consider that the Bank's contribution is only part of the story. The assistance program's outcome is determined by the *joint* impact of four agents: (a) the client; (b) the Bank; (c) partners and other stakeholders; and (d) exogenous forces (e.g., events of nature, international economic shocks, etc.). Under the right circumstances, a negative contribution from any one agent might overwhelm the positive contributions from the other three and lead to an unsatisfactory outcome.

OED measures Bank performance primarily on the basis of contributory actions the Bank directly controlled. Judgments regarding Bank performance typically consider the relevance and implementation of the strategies; the design and supervision of the Bank's lending interventions; the scope, quality, and follow-up of diagnostic work and other AAA activities; the consistency of the Bank's lending with its non-lending work and with its safeguard policies, and the Bank's partnership activities.

Evaluating Assistance Program Outcome

In evaluating the outcome (expected development impact) of an assistance program, OED gauges the extent to which major strategic objectives were relevant and achieved, without any shortcomings. Programs typically express their goals in terms of higher-order objectives, such as poverty reduction. The Country Assistance Strategy (CAS) may also establish intermediate goals, such as improved targeting of social services or promotion of integrated rural development, and specify how they are expected to contribute toward achieving the higher-order objective. OED's task is then to validate whether the intermediate objectives produced satisfactory net benefits, and whether the results chain specified in the CAS was valid. Where causal linkages were not fully specified in the CAS, it is the evaluator's task to reconstruct this causal chain from the available evidence and assess relevance, efficacy, and outcome with reference to the intermediate and higher-order objectives.

Evaluators also assess the degree of client ownership of international development priorities, such as the Millennium Development Goals, Bank corporate advocacy priorities, and safeguards. Ideally, any differences on dealing with these issues would be identified and resolved in the CAS, enabling the evaluator to focus on whether the trade-offs adopted were appropriate. However, in other instances, the strategy may be found to have glossed over certain conflicts or avoided addressing key client development constraints. The consequences could include a diminution of program relevance, a loss of client ownership, or unwelcome side effects such as safeguard violations, all of which must be taken into account in judging program outcome.

To make evaluation less subjective, OED examines a number of elements that contribute to assistance program outcomes. The consistency of evaluation elements is further strengthened by examining the country assistance program across three dimensions:

- A *Products and Services Dimension,* involving a "bottom-up" analysis of major program inputs—loans, AAA, and partnerships
- A *Development Impact Dimension,* involving a "top-down" analysis of the principal program objectives for relevance, efficacy, outcome, sustainability, and institutional impact
- An *Attribution Dimension,* in which the evaluator assigns responsibility for the program outcome to the four categories of actors.

Highly satisfactory

The assistance program achieved at least acceptable progress toward all major relevant objectives and had best practice development impact on one or more of them. No major shortcomings were identified.

Satisfactory

The assistance program achieved acceptable progress toward all major relevant objectives. No best practice achievements or major shortcomings were identified.

Moderately satisfactory

The assistance program achieved acceptable progress toward most of its major relevant objectives. No major shortcomings were identified.

Moderately unsatisfactory

The assistance program did not make acceptable progress toward most of its major relevant objectives, or made acceptable progress on all of them, but either (a) did not take into adequate account a key development constraint or (b) produced a major shortcoming, such as a safeguard violation.

Unsatisfactory

The assistance program did not make acceptable progress toward most of its major relevant objectives, and either (a) did not take into adequate account a key development constraint or (b) produced a major shortcoming, such as a safeguard violation.

Highly unsatisfactory

The assistance program did not make acceptable progress toward any of its major relevant objectives and did not take into adequate account a key development constraint, while also producing at least one major shortcoming, such as a safeguard violation.

Ratings Scale

OED uses six rating categories for **outcome,** ranging from *highly satisfactory* to *highly unsatisfactory.*

Institutional Development Impact

The **institutional development impact** (IDI) of a country assistance program can be rated as: *high, substantial, modest,* or *negligible.* IDI measures the extent to which the program bolstered the client's ability to make more efficient, equitable, and sustainable use of its human, financial, and natural resources. Examples of areas included in judging the institutional development impact of the program are:

- The soundness of economic management

- The structure of the public sector, and, in particular, the civil service
- The institutional soundness of the financial sector
- The soundness of legal, regulatory, and judicial systems
- The extent of monitoring and evaluation systems
- The effectiveness of aid coordination
- The degree of financial accountability
- The extent of building NGO capacity
- The level of social and environmental capital.

Sustainability

Sustainability can be rated as *highly likely, likely, unlikely, highly unlikely,* or, if available information is insufficient, *nonevaluable.* Sustainability measures the resilience to risk of the development benefits of a country assistance program over time, taking into account eight factors:

- Technical resilience
- Financial resilience (including policies on cost recovery)
- Economic resilience
- Social support (including conditions subject to safeguard policies)
- Environmental resilience
- Ownership by governments and other key stakeholders
- Institutional support (including a supportive legal/regulatory framework and organizational and management effectiveness)
- Resilience to exogenous effects, such as international economic shocks or changes in the political and security environments.

ANNEX B: COMPLETED COUNTRY ASSISTANCE EVALUATIONS, FISCAL YEARS 1995–2004

FY95 (1)	FY96 (2)	FY97 (2)	FY98 (10)	FY99[a] (12)	FY00 (10)	FY01 (8)	FY02 (9)	FY03 (8)	FY04 (8)
Ghana	Argentina	Morocco	Albania	Azerbaijan	Argentina	Paraguay	West Bank & Gaza	Peru	Tunisia
	Zambia	Poland	Bangladesh	Cambodia	Burkina Faso	Kazakhstan	Lesotho	Zambia	Bhutan
			Bolivia	Croatia	Cameroon	Morocco	Chile	Eritrea	China
			Côte d'Ivoire	Ecuador	Costa Rica	India	Vietnam	Zimbabwe	Bosnia-Herzegovina
			Kenya	Ethiopia	Egypt	Kenya	Haiti	Lithuania	Armenia
			Malawi	Indonesia	Ghana	Kyrgyz Republic	Bulgaria	Brazil	Moldova
			Mozambique	Jamaica	Papua New Guinea	Mexico	Mongolia	Dominican Republic	Croatia
			Philippines	Maldives	Tanzania	El Salvador	Russia	Jordan	Rwanda
			Thailand	Nepal	Uganda		Guatemala		
			Togo	Sri Lanka	Uruguay				
				Ukraine					
				Yemen					

a. A Country Assistance Note for Honduras was prepared in fiscal 1999 but was converted to a Reach following Hurricane Mitch.

The grey shaded area represents the 25 CAEs that were reviewed to prepare this CAE retrospective.

ANNEX C: SUMMARY OF BOARD/CODE INTERVIEWS

Background and Objectives

This annex is based on an independent research study commissioned to assess the quality and effectiveness of Country Assistance Evaluations (CAE), as perceived by participants in CODE and CODE subcommittee discussions. Primary research objectives included:

- Understanding participants' perception of the most appropriate audiences, objectives and institutional value for CAEs, including the degree to which the documents are perceived to fulfill the purposes for which they were designed
- Identifying potential obstacles to the CAE's overall quality and effectiveness with key audiences
- Assessing the role of CAEs in relation to key objectives (e.g., learning for the future, impact on the development of Country Assistance Strategies, and management accountability)
- Identifying the ways in which the CAE's are used by executive directors (EDs), assistants and advisors, with particular attention to participants' preferences for and evaluations of specific sections and emphases of the document (cover memo, ratings system, counterfactuals, recommendations, and so on)
- Evaluating CAEs in relation to specific criteria (methodological soundness, thoroughness, comprehensiveness, user-friendliness, fairness, objectivity)
- Recommending specific ways to enhance the quality and effectiveness of the documents (for example, timeline for distribution, methodology, style and organization of the documents, input by and dissemination to domestic stakeholders, and the like).

Methodology

To achieve these objectives, the consultant conducted in-depth interviews with 30 former and current CODE Board members, assistants, and advisors. Sixteen of the interviews for the study were conducted in person in Washington, D.C., during late June and early July 2003, with the remaining 14 interviews conducted by telephone between June and August 2003. Each interview lasted about 45 minutes. All participants were promised confidentiality and were provided via e-mail with advanced copies of the interview guidelines.

Participants were recruited from a list of current and former CODE subcommittee attendees provided by OED. The original selection criteria called for the recruitment of current and former EDs, assistants and advisors who had attended at least two CODE subcommittee meetings in the past. Because of limited availability of Board members and their staff during the period when the research was conducted, but the selection criteria were subsequently expanded to include a small number of participants who had attended one CODE subcommittee meeting. This turned out to be a fortuitous decision, since it was discovered during the course of the interviews that reviewing CAEs and preparing for subcommittee discussions and presentations is, in most cases, a collaborative process. EDs and other staff responsible for participating in the CODE discussions or making presentations often rely heavily on summaries and talking points provided by assistants and advisors who do not attend the meetings. Consequently, participants with limited subcommittee attendance were in some cases more familiar with the content and design of the source documents than were EDs and other more frequent attendees. In this light, the sample composition provides a balanced representation of the professional Board staff who utilize the CAEs in preparing for and participating in CODE subcommittee discussions. The final sample included 17 participants who had attended three or more meetings (attendance ranging from 3 to 12); seven participants who attended two meetings each, and six participants who attended one meeting.

Findings

Overall Assessment

There was a broad positive view of CAEs. The majority of those interviewed (24 out of 30) reported that CAEs play an essential role, providing comprehensive, independent assessments to the Board and Bank management that are not available anywhere else. Most participants described the CAE as an indispensable information resource of comparatively high substance and quality that provides thorough assessments of Bank policies and activities. Criticisms of the CAEs were frequently accompanied by sympathetic remarks regarding the inherent difficulties in providing an independent evaluation tool for the Bank and expressions of confidence in OED's commitment to providing independent, objective assessments of Bank performance.

CAE Utilization

Participants indicated that from the Board's perspective, accountability and ensuring that Bank programs are effectively delivered are central concerns, and a credible, independent voice as provided by CAEs was crucial to this process. The considerable learning contribution of CAEs was also noted. Participants are aware of the tension between the accountability and learning functions, and several interviewees stressed the importance of finding the right balance.

Subjects reported that CAEs are routinely used to:

- Prepare for participation in CODE subcommittee meetings
- Prepare for participation in Board meetings, primarily the conclusion and recommendations section of the CAE
- As a background resource for understanding and assessing newly developed Country Assistance Strategies
- As a background resource on specific sector, thematic, or regional issues, to find out how certain issues are treated in specific country settings.

CAEs are used selectively. For the majority of participants, CAEs are not documents to be read from cover to cover but resources from which to pick and choose specific areas of interest. Participants described a typical routine sequence used to review and identify key issues in a CAE: (i) reading the cover memo/executive summary; (ii) looking at the recommendations; (iii) looking at the management response; and then (iv) flipping through the rest of the document to identify specific areas of interest.

Many participants reported that CAEs have a limited impact on country strategy formulation. While a number of participants (8 out of 30) provided firsthand accounts about the positive influence of CAEs on the CAS preparation process, the majority (17 out of 30) expressed doubt that CAEs exert any significant influence on management decisions or policy development. Participants pointed to the frequent lack of formal endorsement of CAEs by the Bank's upper management. This weakened the impact of the CAE at the management and country level, with the result that operational staff were more likely to be dismissive of the findings and recommendations. Management defensiveness was attributed both to an excessive emphasis on formal accountability as well as methodological concerns. All interviewees expressed a desire to find a mechanism by which CAEs could be given more influence.

There is a lack of clarity about how the evaluations can and should be used. A majority of participants (16 out of 30) were not clear on how precisely they might use the CAE in interactions with management or other possible audiences (e.g., government beneficiaries, country teams). They felt that some guidance was needed from OED and suggested that there may be a need for a formal mechanism for integrating CAE recommendations into the CAS.

CAEs were seen as being of limited use to in-country beneficiaries. Factors thought to limit the CAE's usefulness to government leaders and other key beneficiaries include:

- Lack of time and institutional capacity
- Resistance, particularly on the part of government officials, to being evaluated by outsiders
- Over-reliance on Bank priorities and jargon in CAE content
- Lack of coordination/standardization of findings (particularly statistical data) with information provided by other bilateral institutions
- Methodological issues, including lack of input from beneficiaries in the evaluation and lack of up-to-date information that adequately describes the situation in country.

CAEs can be a tool for facilitating interaction between Bank staff and beneficiaries. Some participants felt that the CAE is less a stand-alone resource than a tool for facilitating interactions and feedback between Bank staff and beneficiaries. The process of disseminating, interpreting, and answering questions and concerns raised by a CAE provides an excellent opportunity for cultivating relationships with key constituencies in country—and for identifying and addressing specific country needs (e.g., institutional capacity, donor coordination, etc.). Several participants gave positive examples of their ongoing use of CAEs with country beneficiaries.

Positive CAE Attributes

High-quality analyses. A majority of the participants (23 out of 30) praised the consistently highly quality of the evaluations, with quality defined in terms such as "rich, substantive analysis" or "excellent writing and overall presentation." The most frequently cited caveats were inconsistencies in the findings or a too-positive, pro-Bank style of presentation. Most cautioned that the quality of the evaluations depended on the consultants and staff who prepared them and that there was some variation in quality from one CAE to the next.

Objectivity of findings and assessments. Most participants (21 out of 30) felt that the CAEs are objective; provide a dispassionate, independent assessment of the Bank's performance, and strive to be fair in their assessments and recommendations. Some expressed concern, however, that the text had been massaged to make findings appear less critical for management.

Thoroughness and comprehensiveness. The CAEs were generally described as thorough and comprehensive. A few participants (5 out of 30), however, thought that essential themes (political economy, private sector development, cultural issues) are sometimes overlooked.

Areas Where Improvements are Needed

Inadequate structure and presentation. Many CODE members (16 out of 30) stated that they had difficulty reading and finding relevant information in the CAEs. For most, this is less a problem of length than of organization; to the extent that length meant thoroughness and comprehensiveness, this was considered an asset. It was suggested that the chapter on "Findings and Recommendations," which is most likely to be read by management, be strengthened and moved to the front of the report.

Overly academic focus. A majority of participants (17 out of 30) felt that CAEs should not be a quantitatively focused, academically credible research document, with measurable indicators of past Bank performance, counterfactuals, and the like. Instead, they said, the evaluations should strive to provide a broad, richly textured portrait of the Bank's role in a country's development over time with a strong focus on lessons learned and recommendations that strengthen future Bank strategies.

Excessive attention to operational procedures to the neglect of actual progress on the ground. Many CODE members worried that the documents are sometimes too focused on Bank performance in following certain procedures and achieving certain objectives to the neglect of the impact of the Bank's involvement on the country itself. They said an increased focus on "the situation on the ground" and tracking actual progress would strengthen the CAEs.

Insufficient direct input. Many participants expressed concerns that CAE authors rely too much on secondary research conducted exclusively in the Washington office rather than working directly with country teams, government officials, and other in-country beneficiaries. Most participants specifically acknowledged the inherent difficulties of involving government officials and other country beneficiaries in the evaluation process. Key issues cited include: limited budget and resources at OED, lack of institutional capacity at the country level, potential to confuse or disrupt existing Bank relationships, and potential to limit OED's independence and objectivity. In spite of these challenges and misgivings, the majority of participants insisted that OED should continue to strive for greater country involvement in the CAEs. The Bank's increased emphasis on country-focused strategies (along with related priorities like donor coordination, capacity enhancement and increased reliance on local expertise) have created methodological challenges that can only be addressed with increased participation by country beneficiaries.

Long time frames. Questions were raised about the long time period (10 years +) covered by the evaluations, especially with respect to the issue of Bank and governmental accountability. How can such an evaluation deal with staff turnover on both the management and government sides, unanticipated political and economic changes in the recipient country, and shifts in strategic objectives and priorities in the Bank?

Poor timing of some CAEs. A majority of the participants (17 out of 30) indicated that they sometimes receive the CAEs too late in the strategy development process for them to have a meaningful impact. Participants were sympathetic to the logistical issues that frequently limit the evaluations' availability. However, they would like to receive the CAEs at least two to four weeks before CODE subcommittee meetings, and one to three months before a CAS discussion.

Rating inconsistencies and inaccuracies. The majority of participants regard the ratings as a generally reliable supplement to the information provided in the narrative portion of the CAE report. Some described the ratings as the part of the CAE with the greatest potential for holding management accountable for its actions and ensuring that it pays attention to the findings. Others, however, felt that less attention should be paid to the ratings. Strong concerns were expressed, moreover, that inconsistencies and inaccuracies in the rating system weakened their impact. Most participants indicated that the rating scales need to be sharpened or the relative values more clearly defined for the ratings to be useful.

ANNEX D: BOARD CAS DISCUSSIONS AND CAS DOCUMENTS

Background and Objectives

This review was designed to feed into the CAE retrospective. Its objective is to focus specifically on the utilization of CAEs by the Board of Executive Directors. Transcripts of Board discussions, as well as the relevant CASs and CASPRs, were reviewed in order to examine the possible impact of CAEs on these discussions. The written statements of Board members circulated prior to the discussion were also reviewed. The sample included 13 CAS discussions on the 12 countries for which recent CAEs were available as of April 2003 (one country was discussed twice). The countries were: Bulgaria, Chile, El Salvador, Guatemala, India, Kazakhstan, Kyrgyz Republic, Mexico, Morocco, Peru, Russian Federation, and Vietnam.

Methodology

As a first step, the review looked for explicit references to a CAE in a chairman's contributions to the CAS discussion. Substantial references provided the strongest evidence that the CAE has been read and been used to prepare a chairman's statement. However, even if a chairman did not explicitly refer to the CAE, he might still have read it and been guided in his choice of issues for the CAS discussion. Also, the level of alignment of issues between the CAE and the CAS Board discussion is important information in itself because serious discrepancies may point to underlying problems of analysis and priority setting. It was thus necessary to go beyond counting and categorizing CAE references and to look also at issues raised during the discussion, comparing the subject matter with issues presented in the CAE. All issues raised by chairmen were grouped by frequency and substance and compared with the emphasis they had received in the CAE.

There is a certain risk in giving too much importance to how frequently an issue was mentioned during a Board discussion, because Board members seem to adapt and abbreviate their contributions as the discussion proceeds to avoid redundancy. In many of the transcripts the last few speakers' contributions were much shorter on average, or at least did not contain as many different issues. This may affect how often OED and the CAE are explicitly referred to. However, comments do cluster in certain areas, and this points to issues of particular interest to chairmen in the country context or even overall.

As a basis for the review process, a worksheet was developed for each transcript, listing CAE recommendations and other issues discussed in the relevant CAE. All issues raised by a particular chairman were recorded on the list and further categorized, depending on the substance of the comment and whether it contained an explicit reference to the CAE. If a chairman raised an issue that was not discussed in any detail in the CAE, it was listed in a separate category and the comment assessed as either of superficial or substantial nature. General comments on the CAE without reference to particular issues were put into yet another category. Moreover, explicit references to the CAE were recorded in context, categorized by content, and assessed for guidance aimed at the CAS team.

Findings

Overall Assessment

This review suggests that CAEs have an impact on CAS/CASPR Board discussions, although the impact varies considerably across countries. Utilization of CAEs by Board members overall appears to be modest, with significant variations both between constituencies and between CAEs under discussion. Certain CAEs receive more attention than others, and certain chairmen seem to rely more on the CAEs than others. Overall, there was a good alignment between issues raised in CAEs and subsequent CAS/CASPR discussions, but an explicit linkage to CAEs often cannot be established.

CAE Utilization

About one third of chairmen's statements, either written or oral, explicitly referred to the relevant CAE at least once. In total, there were 85 references to CAEs in the 13 discussions. About a quarter (21) of these were very general references while the remaining 75 percent (64) were issue-oriented. The latter presented a fairly small fraction (6 percent) of the number of issues raised during the Board discussions of these CASs. On the other hand, there was a very close alignment of issues between CAEs and the respective CAS/CASPR Board discussion. This is true whether or not explicit references were made to a CAE and, hence, no clear connection can be established between alignment of issues and use of CAEs by Board members.

The alignment between CAE issues and the points raised by chairmen during the Board discussions is fairly good. For example, one CAS Board discussion came almost a year after the relevant CAE was distributed and hardly anybody referred to it during the discussion. Still, the topics covered in the CAE and the CAS discussion were overall quite well aligned, and CAE recommendations were replicated in several chairmen' comments. The same was true for another discussion that showed particularly good alignment of issues with the CAE, but included no explicit reference to the CAE at all. The evidence suggests that this issue alignment could result from interaction and cooperation between the CAE and the CAS teams in preparation of the CAS.

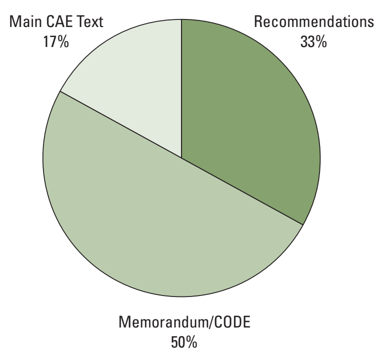
The summary-type sections are referred to most frequently. A closer look at the issue-oriented CAE references suggests selective reading of the CAEs. One third of the references point to the "Recommendations" section of a CAE, and another 50 percent referred to CAE Summaries or to issues raised at a CODE/Subcommittee meeting. Only 17 percent of these references come from the main text (figure 1). This underlines the importance of the summary-type sections of the CAE and of ensuring that the discussion of the CAE with the CODE/Subcommittee is effective.

Some CAEs provoke more discussion than others. A CAE gets more attention when a full CAS is being discussed rather than a CASPR. A CAE gets less attention when the time period elapsed between its distribution and the CAS Board discussion is close to 1 year. To some extent, a CAE with an unsatisfactory outcome rating gets more attention.

Some chairmen make more frequent use of CAEs than others. Ten of the 24 constituencies represented in the Board referred to the CAEs with some regularity. The vast majority of these were either Part I single constituencies or headed by Part I country representatives. Other constituencies seldom or never referred to a CAE.

Some issues are raised more often than others, whether or not they are emphasized in the CAEs under review. Foremost among these is the issue of aid coordination. The discussion is by no means dominated by bilateral aspects of donor coordination but rather is substantive, touching on coordination of the Bank's program with other donor programs, selectivity, collaboration on ESW, harmonization of procedures, and the like. This has implications for future CAEs, as several past CAEs limited their treatment of aid coordination to describing other donors' activities. Other recurring topics of discussion are public sector management, rural poverty, private sector development, and conditionality and lending triggers.

Figure 1 **Distribution of Related References**



ANNEX E: SUMMARY OF BANK STAFF INTERVIEWS

Background and Objectives

This annex is based on an independent research study commissioned to assess the quality and effectiveness of Country Assistance Evaluations (CAEs), as perceived by Bank staff involved in the CAE process. The study focused on the self-reported experiences, perceptions, and expectations of respondents, and is not intended to provide an objective assessment of the quality, presentation, and delivery of CAEs. The aim is to identify key attitudes that inform respondents' receptivity—or lack of receptivity—to the findings and recommendations included in the Country Assistance Evaluations.

Primary research objectives included:

- Understanding participants' perceptions of the most appropriate audiences, objectives and institutional value for the CAEs, including the degree to which the documents are perceived to fulfill the purposes for which they were designed
- Identifying potential obstacles to the CAE's overall quality and effectiveness with key audiences
- Assessing the role of the CAEs in relation to key objectives (e.g., learning for the future, impact on the development of Country Assistance Strategies, and accountability)
- Identifying the ways in which CAEs are used by different audiences in their work
- Evaluating the CAEs in relation to specific criteria (methodological soundness, thoroughness, comprehensiveness, user-friendliness, fairness, objectivity, and the like)
- Recommending specific ways to enhance the quality and effectiveness of the documents (for example, timeline for distribution, methodology, style and organization of the documents, input by and dissemination to domestic stakeholders, and so on).

Methodology

To achieve these objectives, OED commissioned an independent consultant to conduct in-depth interviews with country directors (CDs) and regional vice presidents (RVPs) and focus group interviews with country team members (economists, managers, coordinators). Research was conducted between mid-October and mid-December of 2003. Participants were recruited from lists of current and former Bank staff provided by OED. The research sample for the study included: four current and former RVPs and one Deputy Vice President; five additional staff members from the VPs' offices; 20 current and former country directors; and 18 additional country team Members, five of whom gave in-depth interviews.

All interview participants were promised confidentiality and were provided via e-mail with advance copies of the interview guidelines. Interviews for the study lasted approximately 45 minutes.

Main Findings

Overall Assessment

With a few exceptions, Bank management had broadly negative perceptions of CAEs. Only seven of the 20 country directors interviewed provided a positive assessment of the CAEs and the utility of their findings and recommendations. Two CDs who were familiar with more than one CAE offered a positive assessment of one or more CAEs and a negative assessment of others. Seven CDs had negative views and the perceptions of the remaining four were mixed.

The RVPs interviewed were even less positive than the CDs. Among the key findings from these interviews and focus group discussions was that the CAEs generally received a positive assessment when it was felt that they had had a direct impact on the preparation of a CAS or on other ongoing work, and vice versa.

CD perceptions were not strictly guided by the outcome ratings received for their countries. While there were a few instances of a direct correlation between outcome ratings and assessment of the CAE, the overall picture was different. The seven CDs who provided a positive assessment dealt with three CAEs with positive outcome ratings and four CAEs with negative outcome ratings. The picture was similar for those CDs whose assessment of CAEs was negative. On the other hand, the perceptions of RVPs were primarily determined by their experience with CAEs that received negative ratings. An association between ratings and country team member perceptions in focus groups discussions was difficult to establish, except that many of the criticisms and suggestions were guided by experiences with CAEs with negative outcome ratings.

CAE Utilization

Some CAEs were seen as having a direct impact on the CAS or the Bank's approach in the country. Some participants stated that the CAE for their country had direct impact on strategy and lending and had a noticeable impact on the way they think about and do their work. In one case, it was stated that "in fact, we didn't lend to X for a long time after that." Participants who identified themselves as primary beneficiaries of the CAE were more likely to have interacted directly with the mission team, stressed the positive impact of the evaluations (with and beyond the CAS), encouraged broader dissemination of findings to beneficiaries, and reported consulting CAEs from other countries in their work.

However, many of the CAEs were seen as being of little or no value. Many interviewees revealed limited appreciation and utilization of the CAEs for their countries. Once the formal process and CODE meeting were over, these CAEs tended to be discarded and ignored. Critics said the CAEs: were superficial documents without any real significance for Bank staff work; did not provide any insights or lessons that could change the way the Bank does things; failed to communicate the richness and complexity of the country situation; and did not contain any new ideas that the country team had not already known about.

Positive CAE Attributes

Value as a learning tool. Participants who provided positive assessments of the CAEs focused on the impact of findings on future strategies and operations, even when they were describing critical evaluations of the Bank strategy or the country team. The learning process was seen as being derived more from the informal, ongoing process through which the CAE findings are collected, communicated, and disseminated rather than the CAE as a freestanding, authoritative document.

Areas that Need Improvement

Competence and experience of task managers. An overwhelming majority (16 out of 20) of the CDs and all of the RVPs reported that the competence and experience of the task manager is a key factor in determining their confidence in and receptivity to a CAE. The most commonly cited attributes of a good task manager included: previous operational experience; sensitivity to the pressures and complexities of making decisions and implementing strategies in the field; knowledge of the region and relevant sectors; and a lack of bias.

Degree of interaction with the country team. Of the CDs interviewed for the study, six of the eight participants who provided a positive assessment of one or more CAEs stressed the value and importance of the process through which the CAE was generated. Participants who assessed the CAEs positively were more likely to describe the CAE—and its value for their work—in terms of an ongoing learning process and repeatedly described open and regular communications between the country team and the CAE mission team that had fruitful discussions regarding the key issues to be examined and the appropriate people and resources to be consulted during the evaluation. Conversely, participants who assessed the CAE negatively described a low level of interaction between the country team and the CAE mission team.

Insensitivity to decisionmaking and operational processes. Participants from all groups (including 12 of the country directors, four RVPs, and both of the focus groups) reported that the CAEs frequently fail to pay sufficient attention to the reasoning behind past strategy development or to adequately acknowledge the commitment and professionalism of management and operational staff. The usefulness of even the most positive CAE is sometimes diminished by the evaluation's failure to place decisions within the context in which they were made or to consider the frame of mind of those who developed and implemented the CAS. "I can't believe that our resistance to the findings and the counterfactuals is simply because all of us are that parochial and that resistance to criticism and change," said one respondent. "It's that people's professionalism is not sufficiently acknowledged in the process."

Haphazard CAE timing. Regardless of their assessment of the CAE, participants from all groups agreed that the timing of the evaluations is too haphazard and unpredictable in its current practice to be a useful strategy development tool on an organization-wide basis. Even participants who described ideal experiences with the timing of a CAE indicated that they were aware of instances in which a CAE was made available too late in the process to inform the CAS—an observation which created skepticism about the usefulness of the CAE in their future work. Most participants indicated that, in order to inform the CAS, the CAE would have to be available during the early stages of the strategy development process. Participants' comments regarding the timing of the CAE were closely linked, in many instances, to the degree of interaction with the CAE mission team and the distinction between process and document in understanding the CAE.

Overemphasis on accountability versus learning. Participants who provided positive assessments of the CAEs rarely used the term "accountability." Participants who provided negative assessments, however, repeatedly complained about the CAE's overemphasis on accountability. The most commonly mentioned problem with the CAE's current emphasis on accountability (and one that summarizes most of the other concerns identified) was that it diminishes the value and usefulness at the strategy development and country team level of the lessons and recommendations for the future that the evaluations are meant to provide. Participants from all groups expressed concern that the current focus on accountability both monopolizes the content, methodology, and organization of the CAEs (with key recommendations and findings often buried at the end of the document) and decreases the receptivity of management and country staff to the recommendations that are provided.

For participants from large, complex countries with extensive Bank investment, as well as the RVPs for those respective areas, the credibility of CAEs as accountability measures is further compromised by the failure of the CAE to address the role of the Board and external pressure from G7 countries in directing Bank strategies. This issue—and its impact on the credibility of the CAEs for management and the country team—was mentioned directly by eight of the country directors and three of the RVPs, and was openly discussed in both of the focus groups with country directors.

Insensitivity to damaging country relations. More than half of the participants (including eight CDs, three RVPs and several participants in the focus groups) spoke of the need for damage control after the public dissemination of a negative CAE. For most, damage to the country team's reputation and its ongoing relationships (with governmental and other beneficiaries, other donors, and the general public) is regarded as the key implication of a negative assessment, far more important than the impact on an individual's career or professional development. Participants were much more receptive to negative findings when the CAE team demonstrated sensitivity to the need for understanding and managing the potential impact of a negative evaluation on ongoing operations. In the best cases, participants were granted adequate time to assimilate and prepare (through informal conversations with key beneficiaries) for the distribution of a negative CAE to external audiences.

Weaknesses in the rating system. Participants were also uniformly critical of the rating system used in the CAE. Views were divided between those who believe that the ratings should be ignored or de-emphasized and those who call for a more rigorous approach. Main concerns included:

- Inconsistencies between the ratings and the narrative section of the CAE
- Lack of clarity about the real meanings of the ratings
- Lack of coordination with the quantitative measurements provided by other donor organizations
- Inherent difficulties in applying uniform criteria from one country to another
- General resistance to the concept of graded performance at the country level
- The inability to measure differences across distinct developmental phases (and frustration with OED's practice of combining and averaging ratings from distinct periods)
- Difficulties in attributing country outcomes to Bank performance
- The creation of undue defensiveness among Bank staff.

Weaknesses in methodology and approach. Most participants commented on the lack of a clear, standardized, consistent, and streamlined CAE methodology that can be used effectively anywhere. The focus was seen as too broad to be useful. In place of the CAE's current comprehensive, multi-sector scope, a large number of respondents (8 of the country directors and 2 RVPs) suggested that the evaluations identify a limited number of specific sectors in which a CAE could have the maximum impact. Significant weaknesses were also seen in the counterfactuals that were perceived as too abstract, over-emphasizing the role of the Bank in country development, and too utopian and privileged by hindsight to be useful in making decisions in the future.

Weaknesses in research capabilities. Participants remarked on the excessive reliance on secondary resources and inadequate consultations with country beneficiaries and country team members in preparing CAEs. OED was seen as adopting a "parachute" approach, dropping in and out of the country for as brief a period as possible and doing most of the work from in Washington. Many people were seen as left out on all levels (beneficiaries, other donors, country team, policymakers within the Bank), particularly people who were in key positions at the time when the policies were being made.

ANNEX F: SUMMARY OF COUNTRY EVALUATION METHODS USED BY OTHER AGENCIES

A desk review was undertaken of all publicly available country evaluation methods for each of the major bilateral and multilateral donor organizations. This information, along with a short questionnaire, was sent to the evaluation departments of each of the donor organizations to confirm the findings of the desk review and to request missing information. All evaluation units provided detailed answers to the questionnaire. The European Bank for Reconstruction and Development (EBRD), the International Monetary Fund (IMF), and the United States Agency for International Development (USAID) were dropped from the analysis as they had not yet undertaken any substantial work on country evaluations. This summary reflects information available as of September 2004.

The review showed that about half of the organizations have formal guidelines for carrying out country evaluations and have formats which are similar to OED's. The evaluations have, for instance, chapters on country background, strategy, and assistance provided to the country, followed by chapters on lessons/findings, and then recommendations. As in CAEs, there is a section that deals with other donors' contributions and exogenous factors. The assessments, like CAEs, also follow DAC criteria of relevance, efficiency, effectiveness, and sustainability. Additionally, some include coherence, impact, and institutional development.

Donors differ from OED with regard to ratings, attribution, and the use of counterfactuals. In contrast to OED, other donors do not usually explicitly rate outcomes of their assistance programs (Table F.1). Because of methodological difficulties in dealing with attribution, only three

Table F.1 Use of Key Methodological Tools

Organization	Ratings	Counterfactual	Attribution	Countries Evaluated
Operations Evaluation Department (OED)	■	■	■	70
Multilateral organizations				
African Development Bank (AfDB)		■	■	6
Asian Development Bank (ADB)	a			7
Islamic Development Bank (IDB)		■	■	16
United Nations Development Program (UNDP)				10
United Nations Children's Educational Fund (UNICEF)				7
Bilateral organizations				
Danish Development Agency (DANIDA)				7
(U.K.) Department for International Development (DFID)			■	3
European Union/European Community (EU/EC)				23
Swiss Development Corporation (SDC)				9
Swedish International Development Cooperation Agency (SIDA)				4

a. In the process of developing a rating system.

of the organizations are formally attempting to attribute outcomes of their programs at a country level (AfDB, IDB, and DFID, which is just starting to conduct country evaluations). Most of the others do, however, attempt to establish credible association between inputs and the observed outcomes and impacts. Only two of the organizations reviewed—AfDB and IDB—use counterfactuals in their evaluations. The view of donors was that both attribution and counterfactuals are difficult to conceptualize for Country Assistance Evaluations.

Donors were more concerned about a lack of clear and defined goals and inappropriate or absent monitoring indicators in country strategies than about attribution and counterfactual analysis. The United Nations Children's Educational Fund (UNICEF) is attempting to address this issue by moving to an Integrated Monitoring and Evaluation Plan (IMEP) that is built into the country program at the preparation stage. The IMEP focuses UNICEF programs on results, the use of a logic model to link inputs with results, as well as a monitoring and evaluation (M&E) system to monitor performance. The European Union (EU) has also moved towards constructing a logic diagram for their Country Assistance Strategies.

ATTACHMENT 1: DISCUSSION SUMMARY, CODE SUBCOMMITTEE

On February 28, 2005, the Informal Subcommittee to CODE met to discuss "Country Assistance Evaluation (CAE) Retrospective—OED Self-Evaluation."

Comments by OED. OED representatives noted that the report, which is a self-evaluation, is based on a review of recently completed CAEs, extensive interviews with CODE and CODE subcommittee members, their advisers, and Bank regional staff, as well as assessments by external evaluators. The retrospective addresses two main issues: lessons learned, and ways to improve the CAE instrument. The report found that a country level evaluation provides a more complete picture than a simple aggregation of project outcomes. In preparing CASs, the Bank needs to better understand the country context and pursue development policy lending only when there is strong government ownership. The retrospective also found that the methodological framework could be strengthened by placing more emphasis on the development impact and the outcomes achieved, rather than Bank instruments and compliance issues.

The OED evaluation concluded that future CAEs need to be more results-based, linking the instruments and the outcomes achieved to the Bank's objectives. As for the CAE ratings and interactions with country teams and in-country beneficiaries, OED representatives stressed that it will unbundle the outcome ratings by objective, according to the new framework; and make an effort for more consultations with in-country beneficiaries. They added that if an appropriate rating system for CAS Completion Reports is agreed upon with management, the CAE ratings are likely to be phased out.

Main conclusions and next steps. Members broadly endorsed the conclusions of the CAE retrospective and the suggested revisions of the CAE framework. They stressed the important role of the CAE instrument for shaping the next cycle of country strategies, as well as sharpening the Board discussions of CASs. At the same time, they noted that there is still room for further improvement and emphasized the importance of country ownership, continuous client engagement and better understanding of particular country conditions for producing credible Country Assistance Evaluation reports.

Among the specific issues raised by the Subcommittee were:

Lessons learned. Members agreed that CAEs provide a more complete picture than project level evaluations and stressed the importance of country ownership and understanding the complexities of local political economy in this regard.

Some speakers were concerned that, according to the report, the Board found the CAE instrument more useful than the Bank staff. OED representatives replied that while dialogue is welcome, it is not uncommon for views to differ on a process of evaluation. They added that, despite the differences, there is a growing demand from country units for more OED country-level evaluations.

Some members felt that CAEs should also take into account the effectiveness of development assistance provided by other donors. At the same time, they acknowledged the difficulties in evaluating and distinguishing contributions in joint projects. One speaker was interested in which approach (attribution or contribution) was more appropriate for impact evaluation. OED representatives agreed that attributing individual contributions is a challenging task, but noted that an attempt should be made to document an association between outcomes and the contributions made by the individual donors. Otherwise, one could question the rationale for the individual donor intervention. In the case of completely joint efforts, such as basket funding, attribution could only be determined at the level of the joint effort.

Strengthening the CAE instrument. Members endorsed the revised framework for country evaluations and stressed the crucial importance of the timing of their delivery–ideally 6 to 9 months ahead of CAS presentation. They noted the need to ensure that CASs reflect on the preceding CAEs and that both documents are results-based.

Some members were eager to see conclusions of other instruments, such as Country Policy and Institutional Assessments (CPIAs) and business climate surveys, incorporated into CAEs. OED representatives replied that while the CPIA is a useful tool, it is not necessarily directly related to CAE ratings.

Members underlined the importance of allocating adequate resources for CAE preparation, including involvement of staff with requisite knowledge of country specifics. A member noted that continuous engagement with the client throughout the CAE process will be a significant contribution to strengthening local capacity. OED representatives replied that it uses expert panels and peer reviews to strengthen the quality of its work, and is exploring ways to widen involvement of country teams and the clients at various stages of evaluation.

Risks. One speaker stressed the importance of carefully analyzing the risks (including a thorough look at actual track record and not declarations of intent) before engaging in preparation of a CAS. That might help to improve cost efficiency and redirect resources in situations where country conditions are not adequate for preparing a full-scale CAS. Another speaker noted that some recommendations of the report could lead to a more risk-averse attitude by the country teams, whereas in certain cases, and especially in "turnaround situations," the Bank should be ready to take risks to achieve progress. OED representatives replied that adequate risk mitigation measures would be important to prevent the accumulation of unproductive debt, the burden of which would be ultimately borne by the borrowers, and in turnaround situations, the Bank's financing should be calibrated carefully to progress on implementing reforms.

Donor coordination. Some members felt that the report should have covered the issues of donor coordination, harmonization and the Bank's comparative advantage vis-à-vis other donors. OED representatives replied that its focus is on how effective the Bank is in helping the countries themselves to take leadership, rather than on donor coordination per se, and that OED does not evaluate the assistance provided by other donors. OED has addressed donor coordination in its evaluations of the Poverty Reduction Strategy (PRS) process, the Comprehensive Development Framework (CDF), and a stand-alone evaluation of donor coordination.

Country sample. Some members noted that the sample of CAEs included few small countries or low-income countries under stress (LICUS) and asked whether it might have affected the objectivity of the final conclusions. OED representatives replied that the choice of CAE subjects is often driven by upcoming CASs, thus leaving out many "fragile" and LICUS states which are covered by other instruments, (TSS and ISN). They said OED continues to pay adequate attention to small states and is also preparing an evaluation on LICUS countries.

ENDNOTES

1. See Annex A for a brief description of OED's CAE methodological approach. (This description is also available on OED's website—http://www.worldbank.org/oed/oed_cae_methodology.html.)

2. See Annex B for a list of completed Country Assistance Evaluations.

3. OED has distilled lessons from CAEs in the past. The *Annual Review of Development Effectiveness* (ARDE) distills lessons from the most recent CAEs completed at the time of the ARDE. In 2002, lessons for four African countries were discussed in a workshop organized by the Africa Region. In the same year, OED prepared a note on "Lessons for Low-Income Countries Under Stress" (LICUS) and LICUS-like countries for the Committee on Development Effectiveness. In 2003, lessons from evaluations in transition economies were synthesized for OED's transition economies study.

4. In many of these countries, macroeconomic stabilization was undermined by poor revenue collection, a reflection of weaknesses in tax and customs administration. Financial intermediation suffered because it took time to develop an efficient payment system, and newly privatized banks lacked capacity to make commercial credit decisions. Central banks lack capacity to regulate banks. Private sector development is hampered because the judiciary is often incapable of implementing key legislation for private sector development, including private property rights, creditor rights, bankruptcy legislation, anti-monopoly laws, and the like.

5. The Bank has developed several new instruments—programmatic adjustment loans, adjustable program loans (APLs), and learning and innovation loans (LILs)—to support reforms involving a long learning process, which seem to be well-suited to the situation in some economies.

6. Examples include strengthening of state secretariats of education, implementation of information and evaluation systems for primary education, and the implementation of minimum operational standards in all schools in Brazil; development of a legal and regulatory framework, exploration and development rights, and environmental regulations in Peru; introduction of a performance-based private management contract for water supply in Jordan; strengthening the General Directorate of Roads in Guatemala; capacity improvements in the Ministry of Finance and the central bank in Mongolia; administrative capacity to adjust utilities' tariffs, enforce cash collections, and monitor performance in Russia; and strengthening of the Ministry of Agriculture in Zambia.

7. Bank interventions in Brazil in health, education, and infrastructure were relatively successful because the Government defined primary education, health, and infrastructure as clear development priorities. On the other hand, in Morocco the Government had not yet established clear priorities or an agenda for implementation in education. Two education projects, which closed in recent years, had unsatisfactory outcomes and negligible institutional impact.

8. OED's IDA review and 2003 ARDE found that the link between countries' policy and institutional performance and lending levels has been strengthened. According to an adjustment lending retrospective, in the last few years, most—but not all—adjustment lending has gone to countries with above-average policy performance, for sectors where there was a track record of progress.

9. In Zambia, withholding disbursements until preconditions were being met would likely have forced issues of governance, structural reform, and debt forgiveness to the forefront at an earlier stage. In Zimbabwe, a Bank stance in the 1997–2000 period demanding implementation of reforms rather than mere expressions of good intentions would have sent a strong message to the country's leadership. In Russia, while the Bank's shift to policy-based lending in 1996–97 was to address systemic reform issues, the

message sent to the Russian authorities was that geopolitical considerations would keep the international community's funding window open, regardless of missteps and hesitation in adopting the reform agenda.

10. For example, the Bank's 1993 strategy for Paraguay was too optimistic about the potential for reform and country implementation capacity, given Paraguay's poor track record in these areas. In Zimbabwe, strategies overestimated government receptiveness and willingness to undertake parastatal and civil service reforms. In Haiti, the 1996 strategy did not adequately recognize the risks posed to achievement of objectives from unresolved political and governance issues, which undermined attempts to promote economic development. In Lesotho, despite its experience with the 1993 elections, the Bank was too optimistic in assuming that democratization and stability could be accomplished shortly after the May 1998 elections. As a result, the Bank's assistance strategy did not include contingency plans in the event the democratization process fell apart, as it did.

11. In Kenya, for instance, the appointment of a Change Team in July 1999 and initiation of long-standing economic governance and policy reforms were rewarded with an adjustment loan when conditions for such support as specified in the 1998 country strategy were only partially met. In Zambia, over-optimism led to less results-oriented or vaguely worded conditions in adjustment lending. In Morocco, the Bank provided a policy reform support loan in the late 1990s as a way of rewarding the country's movement toward a more open political system and commitment to reform. The loan was too unfocused to have a major impact on any of the critical reform areas identified in the country strategy. Many of the actions taken prior to Board presentation were first steps, sometimes in the form of studies or plans, and many others did not show concrete results.

12. OED's evaluation of the Heavily-Indebted Poor Country (HIPC) Initiative found that unrealistic growth projections led to debt problems.

13. For an initial attempt to bring about a convergence of aid donor methodologies, see "Best Practice in Country Program Evaluation." OECD DAC Committee on Evaluation, 2004.

14. See OED 2002, page 28.

15. No strategy documents have been prepared for Lesotho, Zimbabwe, or Eritrea.

16. See Annex D for a complete discussion of this review.

17. Annex A contains a full description of the current CAE methodology.

18. See Annex F.

19. Annex A gives a complete description and definition of CAE ratings.

20. "Assistance program" refers to products and services generated in support of the economic development of a client country over a specified period of time, and "client" refers to the country that receives the benefits of that program.

REFERENCES

OECD (Organisation for Economic Co-operation and Development), DAC Committee on Evaluation. 2004. "Best Practice in Country Program Evaluation." Paris.

OED. 2004a. *2003 Annual Review of Development Effectiveness: The Effectiveness of Bank Support for Policy Reform.* Washington, D.C.: World Bank.

———. 2004b. *Economies in Transition: An OED Evaluation of World Bank Assistance.* Washington, D.C.: World Bank.

———. 2003. *Debt Relief for the Poorest: An OED Review of the HIPC Initiative.* Washington, D.C.: World Bank.

———. 2002. *2002 Annual Report on Operations Evaluation.* Washington, D.C.

———. 1999. *1999 Annual Review of Development Effectiveness.* Washington, D.C.: World Bank.