

Influential Evaluations:

Evaluations that Improved Performance and
Impacts of Development Programs

OED
ECD

Copyright © 2004
The International Bank for Reconstruction
and Development/THE WORLD BANK
1818 H Street, N.W.
Washington, D.C. 20433, U.S.A.

All rights reserved.

Manufactured in the United States of America.

The opinions expressed in the report do not necessarily represent the views of the World Bank or its member governments. The World Bank does not guarantee the accuracy of the data included in this publication and accepts no responsibility whatsoever for any consequence of their use.

Acknowledgement

This **Influential Evaluations** study was prepared as a team effort, with substantive contributions from a number of individuals. Michael Bamberger was the principal researcher and advisor for the study, with support from Elaine Ooi (consultant), who also prepared a detailed case study on the China forest evaluation. Other authors of individual case studies, which are presented in the forthcoming companion volume **Influential Evaluations: Detailed Case Studies**, are listed below.

- S.P Pal and Amar Singh, *Improving the Efficiency of the Indian Employment Assurance Scheme*
- Richard Hopkins and Nilanjana Mukherjee, *Assessing the Effectiveness of Water and Sanitation Interventions in Flores, Indonesia*
- Mita Marra, *Broadening the Policy Framework Criteria for Assessing the Viability of Large Dams*
- James Garrett and Yassir Islam, *The Abolition of Wheat-Flour Ration Shops in Pakistan*
- Todor Dimitrov, *Enhancing the Performance of a Major Environmental Project in Bulgaria*

OED would like to thank all of the authors of the evaluation case studies. The summaries presented in this report are the responsibility of the OED team and should not be attributed to the case study authors. The task manager for the **Influential Evaluations** study was Keith Mackay (OEDPK).

A substantive contribution was also made by Dr A. Ravindra, who undertook a detailed review of the impact of the Bangalore citizens' report card; this is presented in a forthcoming OED working paper. Valuable information and feedback on individual case studies were provided by Ananya Basu, Stephen Howes, Jikun Huang, Xu Jintao, Uma Lele, Radhika Nayak, Dr Samuel Paul, Ulrich Schmitt and Susan Shen. The peer reviewers for this paper were Zhengfang Shi and Susan Stout.

Patrick G. Grasso
Acting Manager
Knowledge Programs & Evaluation Capacity Development

Table of Contents

Overview.....	5
The Case Studies	
• Improving the Efficiency of the Indian Employment Assurance Scheme.....	6
• Using Citizen Report Cards to Hold the State to Account in Bangalore, India.....	8
• Assessing the Effectiveness of Water and Sanitation Interventions in Flores, Indonesia.....	10
• Broadening the Policy Framework for Assessing the Viability of Large Dams.....	12
• The Abolition of Wheat-Flour Ration Shops in Pakistan.....	14
• Improving the Delivery of Primary Education Services in Uganda.....	16
• Enhancing the Performance of a Major Environmental Project in Bulgaria.....	18
• Helping Re-assess China’s National Forest Policy.....	20
Designing Useful Evaluations: Lessons Learned.....	22
Additional Resources on Monitoring and Evaluation.....	24

INFLUENTIAL EVALUATIONS OVERVIEW:

EVALUATIONS THAT IMPROVED THE PERFORMANCE AND IMPACTS OF DEVELOPMENT PROGRAMS: CASE STUDIES AND LESSONS LEARNED

PURPOSE

When conducted at the right time, and when they focus on key issues of concern to policy makers and managers, and when the results are presented in a user-friendly format, evaluations can provide a highly cost-effective way to improve the performance and impact of development policies, programs and projects. But evaluations that fail these criteria may produce no useful results—even when they are methodologically sound.

This report presents 8 examples of evaluations that had a significant impact. In many cases it was possible to compare the costs of conducting the evaluation with the economic benefits produced and to show that the evaluation was a highly cost-effective management tool. The cases describe the following evaluations:

- *Improving the Efficiency of the Indian Employment Assurance Scheme*
- *Using Citizen Report Cards to Hold the State to Account in Bangalore, India*
- *Assessing the Effectiveness of Water and Sanitation Interventions in Flores, Indonesia*
- *Broadening the Policy Framework for Assessing the Viability of Large Dams*
- *The Abolition of Wheat-Flour Ration Shops in Pakistan*
- *Improving the Delivery of Primary Education Services in Uganda*
- *Enhancing the Performance of a Major Environmental Project in Bulgaria*
- *Helping Re-assess China's National Forest Policy*

The report concludes with a summary of lessons learned concerning the design of useful evaluations, the extent to which evaluation utilization can be assessed, and the extent to which their cost-effectiveness can be estimated.

A companion volume presents the cases in more detail—**Influential Evaluations: Detailed Case Studies**. It also describes the methodologies used to identify the impacts of the evaluations.

A separate publication—**M&E: Some Tools, Methods and Approaches**—provides a thumbnail sketch of various types of monitoring and evaluation, including several types used in the case studies contained in this *Influential Evaluations* volume. The *M&E Tools* publication describes: their purpose and use; advantages and limitations; costs, skills and time required; and key references. These publications are available from OED's *evaluation capacity development* website: <http://www.worldbank.org/oed/ecd/>

Improving the Efficiency of the Indian Employment Assurance Scheme

The Indian Employment Assurance Scheme (EAS) was launched by the federal government in October 1993 in poor and drought prone districts (blocks) throughout India to assure employment during lean agricultural seasons, and to create economic and community infrastructure to promote sustained employment and development. The scheme had a budget of US\$518 million in 1997-98 and was implemented through the development administration of the state governments under the supervision of the Central Ministry of Rural Areas and Employment (MRAE).

The purpose of the evaluation

The Programme Evaluation Organisation (PEO) was asked by the government's Planning Commission to assess the performance of EAS and to suggest measures for improved performance. In view of reported unsatisfactory performance of EAS and other poverty alleviation schemes, reforms of these schemes were already on the agenda of the government. However, an independent evaluation was needed to judge performance on the basis of hard data from the grassroots, and to determine how exactly the restructuring could be undertaken.

Evaluation methodology

- Review of program records and other secondary sources.
- A multistage stratified sample covered 1120 beneficiaries in 112 villages spread over 14 states.
- Structured interviews were conducted with beneficiaries, community leaders and local and state level officials.
- Qualitative information was obtained from key informants and from direct observation.
- Program records were reviewed to compare actual progress with reported expenditures, projects completed and number of beneficiaries.

Evaluation findings

Program Implementation

- There was little advanced planning in the management of the EAS, and the main concern of local agencies was to spend within a given year as much as possible of the funds allocated to them.
- Local monitoring committees were ineffective and there was considerable misallocation of funds and exaggeration of the number of implemented projects.
- Many villagers were unaware of the details of the scheme.

Utilization of funds

- Administrative delays were a major cause of underutilization of funds.
- Actual fund utilization rates were much lower than reported.

Program Impact

- One fourth of the beneficiaries did not belong to the target group.
- Only around 5% of the target group actually received employment and beneficiaries were employed for less days than reported.

Cost and duration of the evaluation

The evaluation cost approximately \$146,000 and was completed in 15 months.

Recommendations of the evaluation

- All rural employment schemes should be consolidated and integrated with food security schemes.
- More active participation of villagers in selection, implementation and maintenance of assets.
- Stronger role for middle-level government in assessing feasibility of proposed schemes.
- Stricter book-keeping procedures are required.
- Funds should be allocated to the poorest and most needy areas (as identified by a “deprivation/development index”).

Estimating the impacts to which the evaluation contributed

- Consolidation of food security and rural employment programs has increased funds for employment generation by \$1,200 million
- The draw-down of excess public stocks of food grains will save \$72 million (3.6% of EAS budget)
- Post-restructuring of staff will cut about 20% from the wage bill (\$100,000).
- Adherence to the target wage/materials ratio could increase employment under the existing budget by 85%.

The question of attribution: assessing the extent to which the impacts were due to the evaluation

The following methods were used to assess the extent to which the observed impacts can be attributed to the evaluation rather than to other unrelated factors.

- The case study identified the specific references to the evaluation in the planning documents for the restructured Rural Full-Employment Scheme (SGRY)
- The case study identified the specific references to the evaluation in the corresponding chapters of the Mid-Term Appraisal of the Ninth National Development Plan.
- The draft case study was reviewed by social protection and employment specialists from the World Bank New Delhi Office and a number of changes and clarifications were made.

Was the evaluation cost-effective?

While the evaluation was certainly not the only source of information used to reform the program, the evaluation recommendations made a significant contribution to the impacts described above. Even if the evaluation was assumed to have been responsible for only 10% of the impacts it would have produced over \$127 million in benefits for an evaluation cost of around \$146,000.

FOR MORE INFORMATION

- S.P Pal and Amar Singh “Evaluation of the Indian Employment Assurance Scheme”, in OED, **Influential Evaluations: Detailed Case Studies**.

Using Citizen Report Cards to Hold the State to Account in Bangalore, India

In the early 1990s, Bangalore, in common with many other cities in India was experiencing poor quality of delivery of public services such as water, electricity, transport, hospitals and regulation of public land. Most of the population accepted that services would be poor, government would be unresponsive and bribes were the only way to obtain services.

The purpose of the Citizen Report Card (CRC) evaluation

An independent NGO, the Public Affairs Centre (PAC), decided to undertake a Citizen Report Card evaluation. Its purpose was to:

- Solicit and document the views of the users of public services.
- Disseminate widely the findings.
- Use findings to pressure public service providers to improve service quality.

Evaluation methodology

- A survey was administered to a stratified random sample of 1130 households in Bangalore in 1993-4.
- A separate sample of slum dwellers was also covered.
- Respondents provided information on the services they had used during the past six months and all of the agencies with which they had interacted.
- The survey covered telephones, electricity, water and sewerage, public hospitals, transport, public banks and public land regulation.
- The findings were disseminated widely through the mass media, public meetings, and presentations to public service provider agencies.
- The surveys were repeated in 1999 to assess changes since the earlier survey with respect to: overall quality of services, behavior of staff and ease of interaction between ordinary citizens and public service agency staff.

Evaluation findings

The first survey in 1993-4 found:

- Only 10.5 per cent of households were “satisfied” (satisfied plus very satisfied) with services.
- Hospitals, transport and public banks were the only services where satisfaction reached double digits.
- 37.5 per cent of households were “dissatisfied” (dissatisfied plus very dissatisfied) with services.

The follow-up survey in 1999 found:

- The overall percentage of “satisfaction” increased from 10.5 to 40.1.
- The overall percentage of “dissatisfaction” fell from 37.5 to 17.9.
- Improvements were very similar for slum dwellers and all households.
- Public hospitals and electricity showed the greatest improvements. For all services, the proportion of satisfied households increased by at least half.
- There was no reduction in the proportion of households paying bribes.

Cost and duration of the evaluation

Each survey took about 7 months to complete and cost \$10-12,000. In addition, the PAC devoted considerable time to dissemination of the survey findings, persuasion of government departments about the need for change, and direct support for several departments which had asked for assistance.

Recommendations of the evaluation

- The reports included specific recommendations for each agency which were communicated during separate meetings between the PAC and each agency.

General recommendations included:

- Agencies should discuss findings with their staff and agree action plans.
- Measures should be taken to promote systematic feedback from the public.
- Efforts should be made to increase transparency and efficiency, in order to reduce the need and the opportunity for bribes.

Estimating the impacts to which the evaluation contributed

- The CRC raised public awareness of the poor quality of services and encouraged the organization of citizens groups to press for improvements.
- The CRC also catalyzed public service agencies to strengthen their customer orientation and improve the quality of services.
- Households reported no reduction in the level of corruption they faced.
- Similar CRC surveys were initiated in other Indian cities and in other countries as a direct result of the pioneering Bangalore study.

The question of attribution: assessing the extent to which impacts were due to the evaluation

- Extensive discussion of the report card findings in the mass media, and public reporting of follow-up actions, indicate the high visibility of CRC findings.
- A stakeholder survey covering 19 senior municipal agency officials and 5 senior state government officials, representatives of 7 NGOs, and 4 journalists confirmed that the CRC had a catalytic effect on public service agencies and civil society.
- There were significant variations in the commitment of different agencies to the proposed actions and in the long-term changes produced.

Was the evaluation cost-effective?

Available evidence suggests that while other factors were also at work, the CRC made an important contribution to the improvements in public service delivery. The investment of about \$22,000 in the two CRC studies, plus follow-up dissemination and collaboration with government departments, helped contribute to an estimated 50 per cent improvement in satisfaction with all major public services. Thus the report cards do appear to have been highly cost-effective.

FOR MORE INFORMATION

- “Using Citizen Report Cards to Hold the State to Account in Bangalore, India” in OED, **Influential Evaluations: Detailed Case Studies**.
- Samuel Paul. 2002. **Holding the State to Account: Citizen Monitoring in Action**. Books for Change. ACTIONAID, Karnataka, India.
- A. Ravindra, “An Assessment of the Impact of Bangalore Citizen Report Cards on the Performance of the Public Agencies”, OED consultant report, 2003. This report will be published shortly as an OED working paper.

Assessing the Effectiveness of Water and Sanitation Interventions in Flores, Indonesia

In December 1992 Flores Island in eastern Indonesia suffered a major earthquake and tidal wave which claimed several thousand lives and destroyed most of the existing, meager infrastructure. One of the emergency relief efforts was undertaken by the Australian Aid agency (AusAID), and this was later converted into the government/AusAID five-year Flores Water Supply and Sanitation Reconstruction and Development Project (FLOWS). The aim of the FLOWS Project was to promote social and economic development by increasing the provision, access, effective use and sustainability of water supply and sanitation facilities, with emphasis on strengthening of project management.

The purpose of the evaluation

The purpose of the evaluation was to contribute to the new national water and sanitation sector policy, by assessing FLOWS five years after its completion.

Evaluation methodology

- A stratified and geographically representative random sample of 63 sites was drawn from the total of 260 sites which had been covered by the project.
- The Methodology for Participatory Assessment (MPA) was used, with fine-tuning to the local culture and language. This combines participatory research tools with quantitative analysis to assess sustainability and use of water supply and sanitation services, while also assessing the extent of gender and social equity achieved in project processes and outcomes.
- MPA specifically samples marginalized groups such as women and the poor who may not otherwise be consulted.
- Gender-balanced teams of Indonesian researchers, with a mix of technical and social assessment skills, facilitated the participatory assessments.
- The study looked at changes in water supply, sanitation and hygiene conditions through the eyes of both the user community and external researchers, and at the communities' use of the improved facilities.
- Institutional, poverty and gender aspects of project outcomes were studied and their links with service sustainability were investigated.

Evaluation findings

- Water schemes were completed in 87 per cent of villages and most were still working 3-8 years after construction. Almost all toilets still functional.
- 13 per cent of water schemes were never completed, mainly due to unresolved social conflicts between villages.
- Serious drop in service levels at half the sites. 22 per cent of facilities provide little or no water for a quarter of each year.

Did the poor gain access to improved water and hygiene?

- The project design had envisaged and provided only communal water facilities, but many wealthy families also installed house connections.
- The poor had more limited access to toilets and many continued to use open air.
- Project-promoted rules forbidding bathing and washing of babies at communal facilities discouraged better hygiene practices by the poor.
- Decision-making monopolized by wealthier groups.

Financial sustainability was threatened and poor were unfairly burdened

- User fees did not cover recurrent costs or even operational costs.
- Due to flat rate payments the poor paid the same amount for less water and less convenient water than the wealthy.

Cost and duration of the evaluation

The total cost of the evaluation was approximately \$150,000 (\$45,000 for international consultants, and \$105,000 equivalent for the national inputs) and the draft report in English was completed in 12 months.

Recommendations of the evaluation

- Focusing on gender and social equity makes for better management and more sustainable services.
- Give more attention to project mechanisms for translating sector policy into action—deficiencies in policy implementation reduce project performance.
- During the planning phase, more attention should be paid to addressing potential community social conflicts.
- Offering and discussing service and cost options with all sub-groups within communities can minimize later social conflicts.

Estimating the impacts to which the evaluation contributed

- The study reinforced the new national policy by showing that focusing on gender and poverty improves sustainability and effective use of services.
- The findings made policy-makers aware of the challenges of translating policies into practice at the community level.

The question of attribution: assessing the extent to which the impacts were due to the evaluation

- Key evaluation findings were incorporated into the government's new policy document.
- In the stakeholder analysis, both the national planning agency and the donor confirmed the contribution of the study, particularly in identifying gaps between previous policy statements and implementation on the ground.

Was the evaluation cost-effective?

Given that government expenditure in the water and sanitation sector over the next 5 to 10 years might be in the range of \$150 million to \$250 million per annum, the \$150,000 expenditure on the evaluation will have been very cost-effective if the implementation of the evaluation findings could improve program efficiency and sustainability by even a few percentage points. Equity impacts could be even greater if the recommendations can correct revealed weaknesses in gender and poverty targeting of water supply and sanitation.

FOR MORE INFORMATION

- Richard Hopkins and Nilanjana Mukherjee “Assessing the effectiveness of water and sanitation interventions in villages in Flores, Indonesia”, in OED, **Influential Evaluations: Detailed Case Studies**.
- For more detail see “Flores Revisited”, written by Christine van Wijk, Kumala Sari and the Pradipta Paramitha team, Nina Shatifan, Ruth Walujan, Ishani Mukherjee and Richard Hopkins., draft dated December 2002, WSP-EAP.

Broadening the Policy Framework for Assessing the Viability of Large Dams

The debate on large dams in development

In 1993 the World Bank introduced new and broader safeguards for the evaluation of investments in large dams, paying greater attention to social and environmental impacts. The controversy surrounding large dams has made potential borrowers reluctant to approach the World Bank and other development agencies for assistance even for justified projects. Yet many developing countries are unable to finance on their own the scale of investments required to tap fully the economic and social development potential of their river basins and meet ever pressing demands for additional water, power, and flood control. Despite their steady decline as a proportion of World Bank lending, in fiscal 2000 the Bank loaned \$1.064 billion for the construction of large dams.

The purpose of the evaluation

In 1997 the World Bank's Operations Evaluation Department (OED) conducted an evaluation of Bank lending for large dams to:

- Assess whether Bank-financed dam projects had satisfied both social and environmental safeguards current at the time the project was approved and also new safeguards introduced subsequently.
- Identify issues requiring additional research to help further clarify the role of the World Bank.

Evaluation methodology

- Evaluation based on a desk review and survey data collected from a number of borrowing countries and dam management agencies in the field.
- Ex-post, cost-benefit analysis conducted of 50 World Bank-financed large dams constructed between 1956 and 1987.
- All of the dams were initiated before the Bank's current guidelines on involuntary resettlement, dam safety, indigenous people, and environmental protection came into effect.
- Impacts assessed in terms of power generation, irrigation water supply, flood control and navigation, distribution of benefits, and support of poverty alleviation.

Evaluation findings

- 90 per cent of dams met the standards applicable at the time of approval.
- Only one quarter complied with the Bank's current, more demanding safeguard policies.
- Mitigating the dams' adverse social and environmental impacts would have been both feasible and economically justified in 74 per cent of the cases.
- Report not conclusive on whether dams could reach acceptable standards with only minor adjustments or whether substantive re-engineering was required to satisfy environmental, safety and social safeguards.

Cost and duration of the evaluation

The evaluation took two years to complete and cost an estimated \$200,000.

Recommendations of the evaluation

- Strengthen environmental and dam safety safeguards.
- Compensation schemes in cases of involuntary resettlement should be better designed and targeted.

Estimating the impacts to which the evaluation contributed

- The evaluation focused the attention of international stakeholders on a broad set of issues and created a space for debate among all stakeholders.
- The evaluation was also the catalyst for the constitution of the World Commission on Dams which created a mechanism for the integration of both social and economic factors in the evaluation of large dams.
- Within the World Bank the evaluation encouraged greater attention to resettlement plans, environmental assessments and safety issues.
- The evaluation increased attention to minimizing technical and environmental effects of obsolescence and degradation of existing dams.
- The evaluation increased the climate for assessment and accountability within the World Bank, discouraging lending for new dams. A climate of risk aversion, due to the more intensive evaluation of proposed lending, was mentioned as one factor in the decline of lending for large dams. The World Bank now focuses more on dam rehabilitation and safety.

The question of attribution: assessing the extent to which the impacts were due to the evaluation

- The impacts of the evaluation were assessed in a follow-up study which included 18 semi-structured interviews with Bank decision-makers, managers, dam experts and the authors of OED's Large Dams Evaluation.
- Participant observation was used while attending World Bank meetings.
- Bank reports and publications of different stakeholders were also reviewed.
- All of the sources provided a consistent view of the impacts of the evaluation.

Was the evaluation cost-effective?

- Most program managers felt the clear analysis and arguments for and against large dams clarified the issues and provided an analytical framework for decision-making.
- The evaluation was also the main contributing factor in the creation of the World Dams Commission.
- Given the fact that the World Bank alone lent \$1.064 billion for new dams in financial year 2000, and assuming that the evaluation contributed to at least some of the impacts discussed above, the investment of \$200,000 in the evaluation would appear to be highly cost-effective.

FOR MORE INFORMATION

- See Mita Marra "The Large Dams Evaluation" in OED, **Influential Evaluations: Detailed Case Studies**.

The Abolition of Wheat-Flour Ration Shops in Pakistan

Since before Independence the government of Pakistan had been operating a system of wheat-flour ration shops intended to provide subsidized wheat flour to low income groups. By the mid 1980s the system had come under increasing attack because it was inefficient and most of the cheap flour was not in fact reaching the intended target groups.

The purpose of the evaluation

In 1985 the International Food Policy Research Institute (IFPRI) was contracted to conduct an independent evaluation of the wheat flour ration shops. The purpose was to assess their costs and benefits and to recommend whether the system should be discontinued. The government was becoming concerned by widespread corruption of the ration shops and was keen to reduce the government subsidies in an atmosphere of deregulation. However, policy makers were reluctant to address what was perceived as being a very sensitive issue and hoped that a credible and independent outside research institution could provide some support for government action.

Evaluation methodology

Working with local researchers from the Pakistan Institute for Development Economics (PIDE), innovative public opinion polls and household surveys were used to obtain initial data on the availability and use of ration shops. These findings were used to start a dialogue between researchers and Pakistani policymakers about changes to make to the system.

Evaluation findings

- More than 70 per cent of the subsidized wheat-flour never found its way to the ration-shop consumers or subsidized bakeries.
- Very few poor consumers benefited from the subsidies.
- Alternative measures could be taken to reduce the negative impacts of eliminating the program on low-income consumers, the wheat flour shop owners and distributors.

Cost and duration of the evaluation

The study cost around \$500,000 and the first results were communicated to key policy-makers within a year.

Recommendations of the evaluation

- Abolition of the ration shops—this was the report's key recommendation.
- Compensatory measures to compensate low-income consumers for the loss of the subsidy, and ration-shop owners and distributors for the loss of income. It was estimated that the cost of these compensatory measures would be much lower than the cost of the subsidy.

Estimating the impacts to which the evaluation contributed

- The IFPRI evaluation provided reliable evidence which supported the government's decision to abolish the ration shops. Their abolition produced net annual savings to the government of at least \$40 million.
- The rapid, informal communication of the findings before the formal reports were published and at a time when the issue was being debated at the highest policy levels provided solid evidence which allowed the program to be abolished.

The question of attribution: assessing the extent to which the impacts were due to the evaluation

- A follow-up case study was conducted by IFPRI in 1997 to assess the impacts of their research. Key stakeholders were identified and asked what factors shaped their decisions to abolish the ration shops, and the policy space in which they operated. Each was interviewed about the sources of their information and the role of the IFPRI-PIDE research.
- Key policy documents were also reviewed to trace the impacts of the research.
- The interviews showed that policy-makers recognized the important contribution of the research, and that in addition to providing hard evidence, the study by a credible and impartial international agency provided political support, making it easier for policy-makers to approve the difficult and politically sensitive decision to abolish the ration shops.
- Although policy-makers drew upon many sources of information, Pakistani researchers acknowledged that the cold, hard numbers provided by the IFPRI-PIDE research provided "the nail in the coffin" for the ration shops.

Was the evaluation cost-effective?

It is estimated that after taking into account compensatory measures, such as raising the salaries of low-income government workers, the abolition of the ration shops produced net annual savings to the government of around \$40 million. Although it is likely that without the IFPRI studies the decision would ultimately have been made to abolish the ration shops, it seems clear that the IFPRI research contributed significantly to advancing the decision. Using a conservative estimate that the IFPRI research probably advanced the decision by one year, then it contributed to savings of \$40 million. Given that the research only cost \$500,000 the study can be considered very cost-effective.

FOR FURTHER INFORMATION

James Garrett and Yassir Islam, "The Abolition Of Wheat-Flour Ration Shops In Pakistan", in OED, **Influential Evaluations: Detailed Case Studies**.

Improving the Delivery of Primary Education Services in Uganda

In the early 1990s, Uganda like many other developing countries was concerned with the poor performance of public services such as education and health. It was believed that a major cause was the “leakage” of allocated funds which did not reach the frontline agencies, but no research instruments were available to assess the importance of these leakages. In 1996 the World Bank launched an innovative research program in Uganda to track public expenditures and to estimate what proportion of the funds actually reached schools and health facilities. The present case study describes the public expenditure tracking survey (PETS) conducted in the Uganda primary education sector.

The purpose of the evaluation

The purpose of the education PETS was to provide reliable estimates of the proportion of funds allocated by the central government which reached primary schools, and to recommend ways to increase the utilization of approved funds.

Evaluation methodology

The education PETS analyzed the timing of budget flows through various tiers of government and compared budget allocations to actual spending on primary schools. Adequate public accounts were not available to report on actual spending, so surveys were conducted in 250 government primary schools in 19 districts and a panel dataset was created on spending and outputs for 1991–95. The PETS was complemented by a more comprehensive facility-based service-delivery survey, which is not discussed here.

Evaluation findings

Leakage of funds

Only 13 per cent of earmarked funds actually reached schools in 1991-95. The remaining 87 per cent disappeared or was used by district officials for other purposes. About 20 percent of funds allocated for teacher salaries went to “ghost workers” who did not exist or were not working as teachers.

The critical role of parents in funding education

Instead of being stagnant as official statistics indicated, the school survey showed a 60 percent increase in primary enrollments during 1991–95. It was found that primary education was mainly funded by parents, who contributed up to 73 percent of total school spending in 1991. Strikingly, parental contributions continued to increase in real terms despite higher public spending.

The impacts of unequal access to information on public spending

In the absence of central government oversight, local governments and schools bargain over the non-wage education allocations disbursed by central government to local governments. But larger schools were found to receive a larger share of the intended funds (per student), and schools with children of better-off parents also experience a lower degree of leakage. The results suggest that strengthening citizens’ awareness, and their ability to monitor and challenge abuses of the system, are important ways to control corruption.

Cost and duration of the evaluation

The cost of the first education study was around \$60,000. The field surveys took 1-2 months, and the overall study was completed in 5-6 months.

Recommendations of the evaluation

- The findings of the study should be widely disseminated to the public.
- Information on funds approved for, and received by, each school should be widely disseminated through local media and publicly displayed at the school.
- The PETS studies should be repeated periodically to monitor progress.

Estimating the impacts to which the evaluation contributed

- The government began publishing the monthly intergovernmental transfers of public funds through newspapers and radio, and required primary schools to post information on inflows of funds for all to see. This also signaled to local governments that the center had resumed its oversight function.
- Two locally implemented follow-up PETS showed that the flow of non-wage funds improved from 13 percent reaching schools in 1991–95 to about 80 to 90 percent reaching schools in 1999 and 2000.
- Prior to the study most schools did not receive any of these grants, while by 1999 less than 10 per cent of schools were not receiving any of them, and 90 per cent were receiving their full entitlement.
- The study also showed the impact of quantitative data on public services as a tool to mobilize “voice.” While individual complaints can be brushed aside, public feedback backed by systematic comparative data is difficult to ignore and can then provide a spark for public action.

The question of attribution: assessing the extent to which the impacts were due to the evaluation.

- There is a clear “paper trail” of government statements and publications recognizing the importance of the PETS findings.
- As soon as they were received by government the findings of the study were widely disseminated.
- The government of Uganda recently decided to undertake PETS surveys annually in each basic service sector.

Was the PETS study cost-effective?

The best estimate of government annual non-salary education expenditures (capitation grants) for primary education in 1999 is around \$27.7 million. The study estimated the percentage of expenditures reaching schools increased from 13 to at least 80 per cent—an increase of over \$18.5 million. The \$60,000 cost of the PETS study has proved to be highly cost-effective.

FOR MORE INFORMATION

- “Uganda: Improving The Delivery Of Primary Education Services Through Public Expenditure Tracking Surveys” in OED, **Influential Evaluations: Detailed Case Studies**.
- Ritva Reinikka and Jakob Svensson. 2002. **Assessing Frontline Service Delivery**. Development Research Group. World Bank.
www.worldbank.org/wbi/publicfinance/documents/Seco/reinikka_assessing%20frontline.doc

Enhancing the Performance of a Major Environmental Project in Bulgaria

Prior to its privatization, the operations of the Bulgarian KCM metallurgical company had been responsible for widespread hazardous contamination of large residential and agricultural areas. In 2001 the Black Sea Trade and Development Bank (BSTDB) approved a 6-year \$9.2 million loan to KCM to finance a large-scale Environmental Improvement Project promoting the introduction of improved environmental technology and operating methods. The project was also intended to avoid the situation where it would have become necessary to restrict or even terminate operations, thus placing at risk 1,540 jobs in an economically depressed area and 1.3% of Bulgaria's annual exports.

The purpose of the evaluation

The evaluation, which was conducted as a mid-term project review, was intended to assess compliance with the environmental action plan (EAP), so as to avoid potential loss of jobs and fines, and to identify ways in which the efficiency and financial viability of the project could be enhanced.

The evaluation methodology

The evaluation combined a desk review and a 2-day field visit to the borrower's site and a neighboring community. The field visits included: 2 focus groups with the borrower's management and representatives of the local community; 3 semi-structured interviews with the borrower's managers/staff and a key project contractor; and 3 site verifications of compliance with randomly selected components of the project. Unobtrusive observations of safety measures such as smoking restrictions, use of helmets and exposure to and monitoring of toxic gases/substances were also conducted.

To ensure independence and avoid any concern that the evaluation was conducted to justify a pre-decided action such as continuation of a sensitive project, the following procedures were used:

- Clear articulation of the project risks, stakeholders' commitment and external lessons learned.
- Triangulation, i.e. obtaining and comparing sensitive data from at least three independent sources, e.g. Bank, borrower, contractor, press, NGO/community, external auditors, on-site observation/verification.
- Use of the *Good Practice Evaluation Standards* of the Evaluation Cooperation Group of the multilateral development banks.

Evaluation findings

- Due to the borrower's concern with mitigating the effects of falling commodity prices, the implementation of the EAP had not been given sufficient priority and progress was vaguely reported.
- In the effort to mitigate the effect on commodity prices, the borrower breached a hedging covenant which increased KCM vulnerability to volatility in the prices of metals and chemicals.
- Weak enforcement of new safety procedures had exposed KCM to potentially heavy fines for non-compliance with European Union norms.

Cost and duration of the evaluation

The evaluation cost approximately \$4,500 and was completed in 2 months.

Recommendations of the evaluation

- EAP reporting and monitoring systems should be strengthened.
- KCM should adjust its hedging policy in line with industry norms.
- KCM should enhance financial incentives and speed up other measures to ensure enforcement of safety and environmental protection procedures.

Estimating the impacts to which the evaluation contributed

- \$400,000 - \$800,000 additional earnings from advancing the start of zinc production by 3 months.
- \$110,000 - \$220,000 additional earnings from advancing the start of H2S04 production by 10 months.
- \$14,500 - \$29,000 in savings from reduced fines for non-compliance with environmental regulations.
- \$16,500 - \$33,000 in savings from reduced accidents from enforcement of safety regulations.

The question of attribution: assessing the extent to which the impacts were due to the evaluation

The following methods were used to assess the extent to which the observed impacts can be attributed to the evaluation rather than to other unrelated factors.

- A simulation of cause-effect trends was conducted on the basis of with-and-without the evaluation scenarios. The “without” scenario assessed how long it would have taken the necessary information to have reached management and for management to take the decisions, and the likelihood that the decisions would have been taken.
- Extrapolation was used to compare borrower’s estimates from project files, the borrower’s financial department, and estimates on EAP/project timing and impact for each component, and independently verified by the BSTDB environment department.
- A follow-up stakeholder survey with KCM and BSTDB management confirmed that both organizations found the evaluation useful and agreed with and implemented the main recommendations.

Was the evaluation cost-effective?

The evaluation cost \$4,500 and was completed in two months. All of the principal recommendations were accepted and implemented by BSTDB and the borrower, and on the most conservative estimates generated additional revenue and savings of at least \$541,000.

FOR MORE INFORMATION

- Todor Dimitrov, “Enhancing The Performance Of A Major Environmental Project Through A Focused Mid-Term Evaluation: The Kombinat za Cvetni Metali S.A. (KCM) Environmental Improvement Project in Bulgaria” in OED, **Influential Evaluations: Detailed Case Studies**.

Helping Re-Assess China's National Forest Policy

In 1999 the World Bank's Operations Evaluation Department (OED) completed a review of the Bank's 1991 Forest Strategy to assess (i) its impact on overall Bank lending, and (ii) the efficacy of the Bank's role and its impact on forest outcomes. Six country case studies were conducted, including one for China.

Evaluation methodology

- A team of Bank consultants, including a senior Chinese researcher, undertook desk and field work for the China case study evaluation.
- Sixty related projects from the agriculture, transportation and other sectors were examined in addition to the Bank's forest portfolio.
- Field work entailed joint Bank-Chinese project site and household visits, and consultations with forest officials, National Planning Commission and provincial authorities, and other donors. Analysis was supplemented by available empirical work in China and from the Statistics Department.
- The case study also benefited from inputs from a stakeholder's workshop in Beijing, peer review at the World Bank, and a web-based consultation.

Evaluation findings

- The Bank's \$1 billion China forest portfolio, although only a fraction of the country's forest program, had strengthened the government's technical and management capabilities in the sector but was less successful in engaging the government in forest sector policy analysis and dialogue.
- As a result of largely national conservation /afforestation efforts, there had been a 15% increase in forest cover, largely in plantations and shelter belts. Bank lending had contributed additional tree cover of 3.3 million hectares, while diversifying the tree species used, and had helped to increase the involvement of rural poor households in forestry and agroforestry.
- The evaluation also underscored the socioeconomic costs of the government's 1998 logging ban—the government itself had estimated that this would cost at least \$22 billion to redeploy some 2.4 million workers.

Cost and duration of the evaluation

The China evaluation took about 18 months and cost about \$80,000.

Recommendations of the evaluation

- Systematic monitoring, evaluation and policy research should be conducted by China to significantly strengthen its forest policies and programs.
- Forest sector work should be expanded to include agricultural land use changes and the impact of forest policies on farming households.
- While not a formal recommendation, the report pointedly questioned the logging ban and its severe economic consequences for the poor.

Estimating the impacts to which the evaluation contributed

The China case study evaluation contributed directly through:

- Helping legitimize debate among senior officials, researchers and others on forest policy and the government's recently imposed logging ban.
- Engendering broad agreement by the Chinese on the need for improved M&E, and also the need for in-depth research and policy analysis of the impact of forest sector projects on the poor and on biodiversity.
- Fostering participation by the Chinese research community and beneficiary farmers in forest programs, and fostering collaboration among key Chinese stakeholders who had previously not interacted.

The China evaluation also contributed significantly to the creation of the Taskforce on Forests and Grasslands (TFG) of the China Council on Environment and Development, in the context of China's own internal reassessment of its logging ban policy and its emphasis on policy analysis and research. The leader of OED's evaluation of the World Bank's forest sector strategy was invited by the China Council to be a co-chair and leading member of this task force. The TFG, in turn:

- Conducted 1400 household surveys in 10 provinces to elicit grassroots perspectives on the impact of forest programs and policies.
- Documented in a convincing and comprehensive manner the many complex issues affecting the efficacy of the government's forest programs.
- Assessed, through the use of in-depth empirical work, the impacts of major government initiatives in forest conservation and demonstrated some unintended negative impacts.
- Recommended that the government's top-down approach to forest planning and management be replaced by more participatory and flexible approaches.
- Recommended replacement of the complete logging ban with pro-active forest land use planning to achieve sustainable forest management.
- Successfully advocated a strategic approach to M&E in the forest sector.

In response, the government is now revising its forest policy and programs in areas such as forest management and land ownership/use. Collectively, these initiatives constitute a substantive change to the application of the logging ban.

The question of attribution: assessing the extent to which the impacts were due to the evaluation

- The impacts of the OED China case study evaluation and the subsequent work of the Chinese TFG were assessed by a follow-up stakeholder analysis. This was conducted by a Washington-based OED consultant who contacted Chinese and World Bank stakeholders, using primarily telephone and e-mail, to obtain their perspectives on the impacts of the OED study.
- Bank reports and publications of different stakeholders were also reviewed.

Was the evaluation cost-effective?

The OED evaluation and the TFG cost about \$80,000 and \$1.02 million, respectively, and together they have helped to legitimize the need for an in-depth reassessment of China's ban on logging. Given that the government has estimated the redeployment cost alone to be \$22 billion, both evaluation activities can be regarded as potentially highly cost-effective.

FOR MORE INFORMATION

Elaine Ooi, "Re-assessing China's National Forest Policy", in OED, **Influential Evaluations: Detailed Case Studies**.

DESIGNING USEFUL EVALUATIONS: LESSONS LEARNED

Encouraging Utilization

The following factors increase the likelihood that an evaluation will help enhance the performance and impacts of development policies, programs and projects:

- *The importance of a conducive policy environment.* The findings of the evaluation are much more likely to be used if they address current policy concerns and if there is a commitment of key decision-makers to accept the political consequences of implementing the findings.
- *The timing of the evaluation.* The evaluation should be launched when decision-makers have clearly defined information needs. The findings must be delivered in time to affect decisions, and key results must often be communicated informally before the final report is completed.
- *The role of the evaluation.* The evaluation is rarely the only, or even the most important source of information or influence for policy makers and managers. A successful evaluation must adapt to the context within which it will be used, and the evaluator must understand when and how the findings can most effectively be used.
- *Building a relationship with the client and effective communication of the evaluation findings.* It is essential to establish a good relationship with key stakeholders, listen carefully to their needs, understand their perception of the political context and keep them informed of the progress of the evaluation. There should be “no surprises” when the evaluation findings are presented.
- *Who should conduct the evaluation?* The case studies identified two different ways to organize the evaluation, each with advantages and drawbacks:
 - Option 1: The evaluation is conducted by the evaluation unit of the managing or funding agency. This usually has the advantage of better access to the key actors and to data, and a better understanding of the political and organizational context within which the evaluation is conducted. A potential risk, however, is that the evaluator becomes too involved in the political context, losing sight of the “big picture” and finding it difficult to explore sensitive areas.
 - Option 2: The evaluation is conducted by an outside organization or body. This can ensure independence and credibility and can make it easier to explore sensitive issues such as local political pressures or the exclusion of vulnerable groups. However, outside evaluators may have less access to decision-makers and to needed data.
 - A third option, not reflected in these case studies, is to attempt to achieve the advantages of the first two options—by managing and/or conducting an evaluation jointly, involving some combination of external or independent agencies and program staff.

The scope and methodology of the evaluation.

There is no single best evaluation methodology; the approach must be adapted to the specific context, the evaluation questions and priorities, and the available resources. The evaluator will often recommend broadening the proposed scope of the evaluation to assess, for example, the implementation process as well as outcomes, or to study more deeply the social and political context within which the program operates. Most of the evaluations also used a multi-method approach, combining quantitative and qualitative

data collection and analysis methods, both to increase the reliability of the findings and to provide a broader framework for their interpretation.

How much should the evaluation cost?

The value of an evaluation should be assessed, like any other project or program expenditure, in terms of its potential cost-effectiveness. A seemingly “expensive” evaluation may be fully justified if it can produce a reduction in costs or an increase in benefits significantly greater than its own cost. The evaluations described in this volume ranged in cost from as little as \$4,500 to \$500,000, but in each case the question “Was the evaluation justified?” was assessed by comparing the benefits produced with the costs of conducting the study.

While in some cases both costs and benefits can be monetized, in other cases the benefits may concern increased equity, environmental quality or overall program effectiveness—in which case decision-makers must make a judgment as to whether it was “worth” investing in the evaluation to produce a certain set of qualitative benefits.

Assessing utilization of evaluation findings and their cost-effectiveness

Most assessments of evaluations are conducted with time and budget constraints, and often even with political constraints limiting which stakeholders can be contacted and what types of questions can be asked. Within these real-world contexts it is rarely possible to conduct the types of rigorous assessment of evaluation methodology or attribution analysis recommended by the textbooks.

The methodology for assessing utilization will often be limited to a review of available research reports, discussions with the evaluation team (in person or by phone or e-mail) and in some cases soliciting the opinions of local experts (such as staff from the World Bank or other donor resident missions).

Attribution analysis, in which the extent to which the observed policy or program outcomes can be attributed to the effects of the evaluation, is difficult in the best of circumstances (but even more so with the types of constraint listed earlier). The techniques described in the OED publication *M&E: Some Tools, Methods and Approaches* represent a range of analytical tools normally available for attribution analysis. Where possible, the assessment should combine and compare as many of these techniques as are available:

- Construction, where possible of counterfactuals describing the situation which would have obtained if the evaluation had not been conducted (Bulgaria, Pakistan)
- Reviewing references to the evaluation in government reports and planning documents (India, Employment Assurance Scheme; Philippines; Uganda)
- Critical analysis of statements made in the evaluation reports concerning impact of the evaluation (India, Employment Assurance Scheme; Bulgaria; Indonesia; Uganda) and interviews with the evaluators.
- Opinions of stakeholders obtained either in structured interviews (India, Citizen Report Cards) or more commonly in response to e-mail requests (Bulgaria; Indonesia).

In most of the case studies, *cost-effective analysis* was used to compare the financial or numerical benefits of the evaluation with its monetary costs. In most cases it was not difficult to identify the potential impacts of the evaluation (such as reduced administrative costs, increased sales, reduced fines, more beneficiaries), and the challenge was to estimate the proportion of the observed changes which could be attributed to the evaluation. A key lesson is to ensure that the assumptions and methods of estimation of the cost-effectiveness analysis are clearly stated so that the reader can judge the validity of the methods.

Additional Resources on Monitoring and Evaluation

World Wide Web sites

- World Bank independent evaluation:
<http://www.worldbank.org/oed/>
- Strengthening countries monitoring and evaluation systems;
and the papers published as part of this *Influential Evaluations*
study: <http://www.worldbank.org/oed/ecd/>
- Monitoring and Evaluation News:
<http://www.mande.co.uk/news.htm>