# Chapter 4
# Drawing on "Good Practice" Impact Evaluations*

The previous chapters have presented the key methods, issues, and challenges that can arise in evaluating project impact. In reviewing the case studies listed in table 4.1 many illustrative examples emerge from interesting approaches in the design, use of data, choice, and application of analytical methods used, and in-country capacity building. These examples, as well as a discussion of the costs of evaluations and the political economy issues that may arise in implementation, are highlighted below.

The 15 case studies included in the review were chosen from a range of evaluations carried out by the World Bank, other donor agencies, research institutions, and private consulting firms. They were selected as a sample of "good practice" for their methodological rigor, attempting to reflect a range of examples from different sectors and regions. Although each impact evaluation has its strengths and weaknesses, the lessons learned from these experiences should help the project manager or policy analyst intending to design and implement future work.

## Early and Careful Planning of the Evaluation Design

Adequate preparation during the beginning stages of project identification will ensure that the right information is collected and that the findings can be used for mid-course adjustments of project components. With early and careful planning it is possible to incorporate all the elements that contribute to a rigorous impact evaluation, such as a baseline survey with a randomized control group, and qualitative data on the processes that may affect impact.

**Uganda Nutrition and Early Childhood Development Project.** This evaluation, though still not yet under implementation, provides an excellent example of early and careful planning (see World Bank 1998a; Garcia, Alderman, and Rudqvist 1999). The project itself focuses on strengthening the ability of parents and communities to care for children by providing them with knowledge on better childcare practices and by enhancing

* This chapter draws on the best practice case studies in annex I and overview pieces prepared by Gillette Hall and Julia Lane, and Subbarao and others (1999).

opportunities to increase income. It is community-based and implemented by a network of nongovernmental organizations (NGOs). The evaluation component, which was integrated into the project cycle from day one, approaches the ideal in terms of evaluation design. First, it generates baseline and follow-up survey data, along with a randomized control group, so that the program's impact on beneficiary outcomes can be rigorously assessed. Second, it enhances this quantitative component with a participatory (qualitative) monitoring and evaluation (M&E) process.

On the quantitative side, the project was designed to allow for an experimental study design in which parishes will be randomly assigned into treatment and control groups. Health cards will then be used to record data on the child's weight in treatment and control parishes. In addition, the baseline household survey will be conducted before services are delivered to the communities, as well as a follow-up survey of the same households two years later. A rapid review of these data is expected to inform the decision to scale up some components of the intervention during the midterm review of the project. A deeper analysis of the data at the end of the project will guide the design of the second phase of the project.

**Ghana Credit with Education Project.** The evaluation of this project was very complex, with many intermediate steps. The project combines elements of a group lending scheme with education on the basics of health, nutrition, birth timing and spacing, and small business skills. The evaluation generally focuses on assessing the nutritional status of children, women's economic capacity to invest in food and health care, women's knowledge and adoption of breast feeding, and weaning. It begins with a very clear conceptual framework, which is illustrated below. This schematic clearly delineates the inputs, intermediate benefits, and long-term outcomes in a way that both facilitates the development of several models and their interpretation. By carefully planning the evaluation and working with a schematic at an early stage, it was possible to clarify many points in a relatively complex design (see annex 1.6).

## Approaches to Evaluation When There Is No Baseline

In practice, many evaluations do not have adequate data. Evaluations are added after it is possible to do a baseline survey or in the absence of comparison groups. Some examples of this are the Bangladesh Food for Education, Mexico PROBECAT, Czech Active Labor Programs, and Argentina TRABAJAR evaluations. Without a baseline, the controls must be constructed by using the matching methods discussed in the previous chapters. This can, however, be quite tricky. The propensity score match-

*(Text continues on page 71.)*

## Table 4.1 Summary of "Good Practice" Impact Evaluations

| Project: | Country | Database type | Unit of analysis | Outcome measures | Econometric Approach: Randomization | Matching | Reflexive comparisons | Double difference | Instrumental variables | Qualitative | Strengths |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Education** | | | | | | | | | | | |
| Radio Nicaragua | Nicaragua | Baseline and postintervention survey | Students and classrooms | Test scores | Yes | No | Yes | No | No | No | Questionnaire design |
| School Autonomy Reform | Nicaragua | Panel survey and qualitative assessments | Students, parents, teachers, directors | Test scores, degree of local decision-making | No | Yes | Yes | No | No | Yes | Qualitative-quantitative mx |
| Textbooks | Kenya | Baseline and postinterventions urvey | Students, classrooms, teachers | Test scores | Yes | No | Yes | Yes | Yes | No | Analysis of confounding factors |
| Dropping out | Philippines | Baseline and postintervention survey | Students, classrooms, teachers | Test scores and dropout status | Yes | No | Yes | Yes | Yes | No | Cost-benefit analysis; capacity building |

*(Table continues on the following page.)*

**Table 4.1** *(continued)*

| *Project:* | *Country* | *Database type* | *Unit of analysis* | *Outcome measures* | *Econometric Approach* | | | | | | *Strengths* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | *Randomization* | *Matching* | *Reflexive comparisons* | *Double difference* | *Instrumental variables* | *Qualitative* | |
| **Labor Programs** | | | | | | | | | | | |
| TRABAJAR | Argentina | Household survey, census, administrative records, social assessments | Workers, households | Income, targeting, costs | No | Yes | No | No | Yes | Yes | Judicious use of existing data sources, innovative analytic techniques |
| PROBECAT | Mexico | Retrospective and labor force surveys | Workers | Earnings and employment outcomes | No | Yes | No | No | No | No | Matching technique |
| Active Labor Programs | Czech Republic | Retrospective mail surveys | Workers | Earnings and employment outcomes | No | Yes | No | No | No | No | Matching technique |
| **Finance** | | | | | | | | | | | |
| Micro Finance | Bangladesh | Postintervention survey plus administrative records | Households | Consumption and education | Yes | Yes | No | Yes | No | No | Analysis of confounding factors |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Credit with Education | Ghana | Baseline and post-intervention survey | Mother-and-child pairs | Income, health, and empowerment | Yes | Yes | Yes | No | No | Yes | Use of qualitative and quantitative information |
| Health Financing | Niger | Baseline and postintervention survey plus administrative records | Households and health centers | Cost recovery and access | No | Yes (on districts) | Yes | No | No | No | Use of administrative data |
| **Food and Nutrition** | | | | | | | | | | | |
| Food for Education | Bangladesh | Household expenditure survey | Households and communities | School attendance | No | No | No | No | Yes | No | Imaginative use of instruments to address selection problem with standard data |
| Health, Education, and Nutrition | Mexico | Baseline and post-intervention surveys | Households | Health, education, and nutrition outcomes | Yes | Yes | Yes | Not known | Not known | No | Clear conceptualization; analysis of confounding factors |
| **Infrastructure** | | | | | | | | | | | |
| Social Investment Fund | Bolivia | Baseline and follow-up surveys | Households, projects | Education and health indicators | Yes | Yes | Yes | Yes | Yes | No | Range of evaluation methodologies applied |

*(Table continues on the following page.)*

**Table 4.1** *(continued)*

| *Project:* | *Country* | *Database type* | *Unit of analysis* | *Outcome measures* | *Econometric Approach* | | | | | | *Strengths* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | *Randomization* | *Matching* | *Reflexive comparisons* | *Double difference* | *Instrumental variables* | *Qualitative* | |
| Rural Roads | Vietnam | Baseline and follow-up surveys | Households, communities | Welfare indicators at household and commune levels | No | Yes | Yes | Yes | Yes | No | Measures welfare outcomes |
| **Agriculture** | | | | | | | | | | | |
| National Extension Project | Kenya | Panel data, beneficiary assessments | Households, farms | Farm productivity and efficiency | No | No | Yes | No | No | No | Policy relevance of results |

ing technique used in the Argentina TRABAJAR project to construct a control group with cross-sectional data on program participants and nonparticipants provides a good example.

**The TRABAJAR II Project in Argentina.** This project was focused on providing employment at low wages in small social and economic infrastructure subprojects selected by community groups. The impact evaluation of the program was designed to assess whether the incomes of program participants were higher than they would have been had the program not been in place. The most commonly used methods to estimate household income without the intervention were not feasible in the case of the TRABAJAR program: no randomization had taken place to construct a control group to use in comparing the income of project beneficiaries; and no baseline survey was available, ruling out the possibility of conducting a before-and-after evaluation.

The TRABAJAR evaluation instead used existing data to construct a comparison group by matching program participants to nonparticipants from the national population over a set of socioeconomic variables such as schooling, gender, housing, subjective perceptions of welfare, and membership in political parties and neighborhood associations by using a technique called propensity scoring. The study demonstrates resourceful use of existing national household survey data—(the Encuesta de Desarrollo Social (EDS)—in generating the comparison group, combined with a smaller survey of TRABAJAR participants conducted specifically for the purposes of the evaluation. The smaller survey was carefully designed so that it used the same questionnaire as the EDS and the same interview teams and was conducted at approximately the same time in order to successfully conduct the matching exercise. This technique was possible in the TRABAJAR case because a national household survey was being canvassed and the evaluators could take advantage of this survey to oversample TRABAJAR participants. The same interview teams were used for both the national and project surveys, resulting in efficiency gains in data collection (see annex 1.1).

**Czech Labor Market Programs Evaluation.** This evaluation attempted to cover five active labor programs to (a) determine whether participants in the different programs were more successful in reentering the labor market than were nonparticipants and whether this varied across subgroups and with labor market conditions; and (b) determine the cost-effectiveness of each program and make suggestions for improvements. The evaluation used a matching technique because no baseline data were collected. The evaluators surveyed participants and then chose a random sample of nonparticipants. Since the nonparticipants were systematically older and less

educated, the evaluators needed to construct a reasonable comparison group for each program. This was done by taking each participant in turn and comparing them to each individual in the nonparticipant pool on the basis of seven characteristics: age, gender, education, number of months employed prior to registration, town size, marital status, and last employment type. The closest match was then put into the comparison group. Although this approach is straightforward, there is the potential for selection bias—that the nonparticipant group is systematically different from the participant group on the basis of unobservables (annex 1.5).

## Dealing with Constraints on Developing Good Controls

At times, randomization or experimental controls are possible but not politically feasible. In this case, the randomization can be carried out by taking advantage of any plans to pilot the project in certain restricted areas. Areas in which the project will be piloted can initially be randomly selected, with future potential project areas as controls. Over time, additional communities can be randomly included in the project. Three examples illustrate how to handle a situation in which randomization was politically or otherwise infeasible. In Vietnam, a rural transport project will be evaluated with limited information and no randomization. The Honduras Social Investment Fund provides an example of how to construct a control group in demand-driven projects, using an ex post matched comparison based on a single cross-section of data. Evaluating demand-driven projects can be particularly difficult given that it is not known which projects or communities will participate in the project ahead of time. And third, the evaluation of the Bolivian Social Investment Fund in the Chaco Region provides a good example of how to incorporate randomization in demand-driven projects in a way that allows targeting.

**The Vietnam Rural Roads Project.** This project aims at reducing poverty in rural areas by improving access to rural communities and linking them to the district and provincial road networks. The design of the impact evaluation centers on baseline and follow-up survey data collected for a sample of project and comparison-group communities identified through matched-comparison techniques. Baseline and postintervention information on indicators such as commune-level agricultural production yields, income source diversification, employment opportunities, availability of goods, services and facilities, and asset wealth and distribution will be collected from a random sample of project (treatment) and nonproject (comparison) communes. These data will be used to compare the change in outcomes before and after the intervention between project and nonproject communes using "double differencing."

Ideally, treatment and comparison communes should be equivalent in all their observed and unobserved characteristics, the only difference between them being that treatment communes benefit from the intervention whereas comparison communes do not. Since random assignment to treatment and comparison groups had not taken place, and the requisite data to make informed choices on appropriate controls were not available at the time of sample selection, random samples of project communes and nonproject communes were drawn. Specifically, project communes were selected from a list of all communes with proposed projects in each province. Next, comparison communes were selected from a list of all remaining communes without proposed projects but in the same districts as treatment communes. Using information collected for the evaluation, propensity score matching techniques will then be used to ensure that selected nonproject communes are appropriate comparison groups. Any controls with unusual attributes relative to the treatment communes will be removed from the sample (annex 1.15).

**Honduran Social Investment Fund.** The Honduran Social Investment Fund (FHIS) (see World Bank 1998b) aims to improve the living conditions for marginal social groups by financing small-scale social and economic infrastructure subprojects. The FHIS is a demand-driven institution that responds to initiatives from municipalities, government ministries, NGOs, and community groups by providing financing for investments in infrastructure, equipment, and training. The impact evaluation of the FHIS uses matched-comparison techniques, drawing the treatment group sample randomly from a list of communities in which FHIS projects have been in operation for at least one year. The comparison group, by contrast, was selected from a list of "pipeline" projects—those that have been requested and approved but for which the FHIS investment has not yet taken place. In theory, comparison-group communities are automatically matched to project communities according to the self-selection process and FHIS project approval criteria. A household survey was then conducted in both treatment and comparison-group communities, complemented by a qualitative evaluation component (focus groups and interviews with key informants) conducted in a subset of treatment communities. This initial evaluation is a first step toward establishing an ongoing M&E system that will be eventually integrated into FHIS operations. In particular, the data collected from communities with pipeline projects will become a useful baseline from which to track future changes in impact indicators, after FHIS investment takes place.

**Educational Investments in the Chaco Region of Bolivia.** Education projects financed by the Bolivian Social Investment Fund (SIF) are aimed

at upgrading physical facilities and training teachers in rural public school. Delays in the implementation of the project in the Chaco Region and limited funds for school upgrading provided an opportunity to use an experimental evaluation design while also ensuring that the neediest schools benefit from the project. Schools in the Chaco Region were ranked according to a school quality index based on the sum of five school infrastructure and equipment indicators: electric lights, sewerage, water source, desks per student, and square meters of space per student. Only schools below a particular cutoff value were eligible for a SIF intervention. Among eligible facilities, the worst-off schools were automatically selected to benefit from investments financed by SIF. The next highest priority group contained 120 schools, but funds were available to upgrade only less than half of them. Thus, eligible schools in this second priority group were randomly assigned to treatment or comparison groups, providing the conditions for an experimental evaluation design (Annex 1.4).

## Combining Methods

For most evaluations, more than one technique is required to achieve robust results that address several evaluation questions. Each question may necessitate different techniques, even within one project design. Three examples illustrate how several techniques were combined in one evaluation; the Bolivia Social Fund, the TRABAJAR Evaluation in Argentina, and the Nicaragua School Reform.

**The Bolivia Social Fund.** Social funds generally include several different types of subprojects, and thus designing an evaluation can involve several approaches. In the Bolivia Social fund, the pattern of project implementation dictated evaluation methods. In the case of education, schools that were to receive the intervention had already been identified; therefore randomization could not be used. Instead, matching methods were adopted. In the case of health projects, reflexive methods were used because the intervention was to be implemented in all health centers in the region (see Annex 1.4).

**Using a Broad Mix of Evaluation Components—Argentina TRABAJAR II.** The TRABAJAR evaluation includes an array of components designed to assess how well the program is achieving its policy objectives. The first component draws on household survey data to assess the income gains to TRABAJAR participants. The second component monitors the program's funding allocation (targeting), tracking changes over time as a result of reform. This component generates twice-yearly feed-

back used to refine program targeting. Additional evaluation components include a cost-benefit analysis of infrastructure projects, and social assessments designed to provide community feedback on project implementation. Each of these components has been conducted twice. Three future components are planned. The matched-comparison research technique will be applied again to assess the impact of TRABAJAR program participation on labor market activity. Infrastructure project quality will be reassessed, this time for projects that have been completed for at least one year to evaluate durability, maintenance, and utilization rates. Finally, a qualitative research component will investigate program operations and procedures by interviewing staff members in agencies that sponsor projects as well as program beneficiaries.

The evaluation results provide clear direction to policy reform. The first evaluation component reveals that the TRABAJAR program is highly successful at targeting the poor—self-selection of participants by offering low wages is a strategy that works in Argentina, and participants do experience income gains as a result of participation. The second component finds that the geographic allocation of program funding has improved over time—the program is now more successful at directing funds to poor areas; however, the ongoing evaluation process indicates that performance varies and is persistently weak in a few provinces, findings to which further policy attention is currently being directed. Disappointing evaluation results on infrastructure project quality have generated tremendous efforts by the project team at improving performance in this area through policy reform—insisting on more site visits for evaluation and supervision, penalizing agencies with poor performance at project completion, and strengthening the evaluation manual. And finally, the social assessments uncovered a need for better technical assistance to NGOs and rural municipalities during project preparation and implementation, as well as greater publicity and transparency of information about the TRABAJAR program (Annex 1.1).

**Nicaragua's School Autonomy Reform.** In 1993, the Nicaraguan Government took decisive steps to implement a major decentralization initiative in the education sector granting management and budgetary autonomy to selected primary and secondary schools. The goal of the reforms is to enhance student learning: as school management becomes more democratic and participatory, local school management and spending patterns can be allocated toward efforts that directly improve pedagogy and boost student achievement. The impact of this reform has been evaluated by using a combination of quantitative and qualitative techniques to asses the outcome as well as the process of decentralization. The purpose of the qualitative component is to illuminate whether or not the

intended management and financing reforms are actually taking place in schools, and why or why not. The quantitative component fleshes out these results by answering the question, "Do changes in school management and financing actually produce better learning outcomes for children?" The qualitative results show that successful implementation of the reforms depends largely on school context and environment (that is, poverty level of the community), whereas the quantitative results indicate that increased decisionmaking by schools is in fact significantly associated with improved student performance.

Different but complementary methodologies and data sources were used to combine both approaches. On the one hand, the quantitative evaluation followed a quasi-experimental design in which test scores from a sample of students in autonomous schools (treatment group) are compared with results from a matched sample of nonautonomous public schools and private schools (comparison group). Data for this component of the evaluation were collected from a panel of two matched school-household surveys and student achievement test scores. The qualitative evaluation design, on the other hand, consisted of a series of key informant interviews and focus group discussions with different school-based staff and parents in a subsample of the autonomous and traditional schools included in the quantitative study.

Using both qualitative and quantitative research techniques generated a valuable combination of useful, policy-relevant results. The quantitative work provided a broad, statistically valid overview of school conditions and outcomes; the qualitative work enhanced these results with insight into why some expected outcomes of the reform program had been successful while others had failed and hence helped guide policy adjustments. Furthermore, because it is more intuitive, the qualitative work was more accessible and therefore interesting to Ministry staff, which in turn facilitated rapid capacity building and credibility for the evaluation process within the ministry (Annex 1.11).

## Exploiting Existing Data Sources

Existing data sources such as a national household survey, census, program administrative record, or municipal data can often provide valuable input to evaluation efforts. Drawing on existing sources reduces the need for costly data collection for the sole purpose of evaluation, as illustrated in the case of the Vietnam Rural Roads evaluation. Furthermore, although existing data may not contain all of the information one would ideally collect for purposes of the evaluation, innovative evaluation techniques can often compensate for missing data, as shown in the Kenya National Agricultural Extension Project.

**The Vietnam Rural Roads Project.** The data used in this evaluation draw on an effective mix of existing national and local data sources with surveys conducted specifically for the purposes of the evaluation. The evaluation household survey is efficiently designed to replicate a number of questions in the Vietnam Living Standards Survey so that, drawing on information common to both surveys, regression techniques can be used to estimate each household's position in the national distribution of welfare.

The evaluation draws extensively on commune-level data collected annually by the communes covering demographics, land use, and production activities. This data source is augmented with a commune-level survey conducted for the purposes of the evaluation. Two additional databases were set up using existing information. An extensive province-level database was established to help understand the selection of the provinces into the project. This database covers all of Vietnam's provinces and has data on a wide number of socioeconomic variables. Finally, a project-level database for each of the project areas surveyed was also constructed in order to control for the magnitude of the project and the method of implementation in assessing project impact (Annex 1.15).

**The Kenya National Extension Project (NEP).** The performance of the Kenya National Extension Project (NEP) has been controversial and is part of the larger debate on the cost-effectiveness of the training and visit (T&V) approach to agricultural extension services. In the Kenyan context, the debate has been elevated by, on the one hand, very high estimated returns to T&V reported in one study (Bindlish, Evenson, and Gbetibouo 1993, 1997) and, on the other, the lack of convincing visible results, including the poor performance of Kenyan agriculture in recent years.

The disagreement over the performance of NEP has persisted pending the results of this evaluation, which was designed to take a rigorous, empirical approach to assessing the program's institutional development and impact on agricultural performance. The evaluation uses a mix of qualitative and quantitative methods to ask highly policy-relevant questions and reveals serious weaknesses in the program: (a) The institutional development of NEP has been limited, and after 15 years there is little improvement in the effectiveness of its services; (b) the quality and quantity of service provision are poor; and (c) extension services have only a small positive impact on farm efficiency and none on farm productivity.

The evaluation is able to draw an array of concrete policy conclusions from these results, many of which are relevant to the design of future agricultural extension projects. First, the evaluation reveals a need to enhance T&V targeting, focusing on areas and groups where the impact is likely to be greatest. Furthermore, advice needs to be carefully tailored to meet

farmer demands, taking into account variations in local technological and economic conditions. Successfully achieving this level of service targeting calls for appropriate flows of timely and reliable information—hence a program M&E system generating a constant flow of feedback from beneficiaries on service content. In order to raise program efficiency, a leaner and less intense T&V presence with wider coverage is likely to be more cost-effective. The program's blanket approach to service delivery, using a single methodology (farm visits) to deliver standard messages, also limits program efficiency. Institutional reform is likely to enhance the effectiveness of service delivery. Decentralization of program design, including participatory mechanisms that give voice to the farmer (such as cost sharing and farmer organizations) should become an integral part of the delivery mechanism. Finally, cost recovery, even if only partial, would provide appropriate incentives, address issues of accountability and quality control, make the service more demand-driven and responsive, and provide some budgetary improvement (Annex 1.8).

## Costs and Financing

There are no doubt many costs involved in carrying out an impact evaluation, which explains why some countries are reluctant to finance such studies. These costs include data collection and the value of staff time for all the members of the evaluation team. Financing for an impact evaluation can come from within a project, other government resources, a research grant, or an outside donor. Information for a sample of World Bank evaluations shows that although many countries assumed the majority of the evaluation costs, the successful implementation of an impact evaluation required substantial outside resources beyond those provided for in a project's loan or credit. These resources came from a combination of the following sources: (a) a World Bank loan or credit financing for the data collection and processing; (b) the government, through the salaries paid to local staff assigned to the evaluation effort (as explained in table 4.1, these staff costs have not been included in the calculation of the evaluation costs conducted for the cases reviewed here because of data limitations); (c) World Bank research grants and bilateral donor grants that financed technical assistance from consultants with specific expertise required for the evaluation; and (d) the World Bank overhead budget through the staff time provided to guide the impact evaluation and often actively participate in the analytical work.

Although few impact evaluations document the cost of carrying out the work, table 4.2 provides cost estimates for a sample of impact evaluations with World Bank involvement. These cost estimates do not, however, include the value of the staff time contributed by client country

**Table 4.2 Summary of Estimated Costs from Several World Bank Impact Evaluations**

| Project | Estimated cost of evaluation ($)[a] | Cost as % of total project cost[b] | Cost as % of IBRD loan or IDA credit[b] | Breakdown of Evaluation Costs (%) | | | |
|---|---|---|---|---|---|---|---|
| | | | | Travel[c] | World Bank staff | Consultants | Data collection |
| Nicaragua School-Based Management | 495,000 | 1.26 | 1.5 | 8.1 | 18.1 | 39.0 | 34.8 |
| El Salvador School-Based Management | 443,000 | 0.60 | 1.3 | 7.7 | 7.4 | 25.8 | 59.2 |
| Colombia Voucher Program | 266,000 | 0.20 | 0.3 | 9.4 | 9.8 | 21.8 | 59.0 |
| Honduras Social Fund | 263,000 | 0.23 | 0.9 | 3.0 | 11.5 | 53.2 | 32.3 |
| Nicaragua Social Fund | 449,000 | 0.30 | 0.8 | 4.9 | 33.0 | 7.8 | 55.7 |
| Bolivia Social Fund | 878,000 | 0.50 | 1.4 | 3.4 | 14.6 | 12.9 | 69.1 |
| Trinidad and Tobago Youth Training | 238,000 | 0.80 | 1.2 | 7.6 | 11.5 | 17.9 | 63.1 |
| Average | 433,000 | 0.56 | 1.0 | 6.3 | 15.1 | 25.5 | 53.3 |

a. This cost does not include the cost of local counterpart teams not financed from the loan or credit. The figures refer to the time period under which the projects in the evaluation sample were selected, not total financing ever provided by the Bank and others to those institutions.
b. These costs as a percentage of the loan or credit or of the project are presented as a reference only. In many cases the actual financing for the evaluation was obtained from sources outside of the project financing.
c. The travel cost estimates include mission travel for World Bank staff and international consultants to the client countries, as well as travel from client country counterparts, particularly to participate in strategy sessions and analytical workshops with international consultants and World Bank staff.
*Source:* World Bank Project Files.

counterparts (which can be significant) because this information was unavailable. As a benchmark, in the 8 cases above it was not unusual to have up to five staff assigned to the evaluation effort for several years, a level of effort sufficient to substantially raise the cost of the evaluation in many of the cases.

The average estimated cost for the impact evaluation was $433,000. This reflects a range from $263,000 for the evaluation of a vocational skills training program for unemployed youth in Trinidad and Tobago to $878,000 for the evaluation of the Bolivian Social Investment Fund. Spending on the impact evaluations for the projects below reflects, on average, 0.6 percent of the total cost of the project (which sometimes included financing from several donors), or 1.3 percent of the cost of the International Bank for Reconstruction and Development (IBRD) loan or the International Development Association (IDA) credit. The most expensive components of the evaluations listed below were data collection and consultants, both local and international. In many of the cases travel costs included local staff travel to meet with World Bank staff and researchers for strategy sessions and training because capacity building for client country staff was a key objective. The two examples below for the impact evaluations of projects in Trinidad and Tobago and Bolivia illustrate some of the variation that can arise in program costs.

The vocational skills training program evaluation in Trinidad and Tobago took advantage of a national income and employment survey to oversample program graduates and create a comparison group from a subset of the national sample. In addition, the evaluation team helped design and use available administrative data from records of program applicants, so preintervention data were available and no enumeration was required to locate program graduates. The total sample size for each of the three tracer studies was approximately 2,500 young people, counting both the treatment and comparison groups. There was only one short questionnaire administered in the survey, and the questionnaire was given only to the program graduates. Trinidad and Tobago is a small country, communities are relatively easy to access by road, and English is the common language in the country and among program graduates.

The Bolivia Social Fund (SIF) evaluation used its own baseline and follow-up surveys of treatment and comparison groups to evaluate interventions in health, education, water, and sanitation. There were no national surveys available from which to conduct analyses or carry out oversampling, which placed the entire burden of data collection on the evaluation. The sample of treatment and comparison groups consisted of close to 7,000 households and 300 facilities interviewed in both the 1993 baseline survey and 1998 follow-up survey.

In Bolivia, the data collection instruments for the impact evaluation consisted of portable laboratories for conducting laboratory-based water quality tests, achievement tests, and eight questionnaires for informants from households and facilities. The eight questionnaires consisted of two household questionnaires (one for the principal informant and one for women of childbearing age), a community questionnaire, four different

health center questionnaires for the different types of health centers (ranging from small community clinics to hospitals), and a school questionnaire for the school director. To assess targeting, the evaluation included a consumption-based measure of poverty, which required the collection of detailed consumption data from households as well as regional price data from communities. The fieldwork was conducted in rural areas where the majority of the SIF projects are located and included a random sample of rural households that were often accessible only by foot or on horseback. Finally, the questionnaires had to be developed and administered in Spanish, Quechua, and Aymara.

## Political Economy Issues

There are several issues of political economy that affect not only whether an evaluation is carried out but also how it is implemented. The decision to proceed with an evaluation is very much contingent on strong political support. Many governments do not see the value of evaluating projects and thus do not want to invest resources in this. Additionally, there may be reluctance to allow an independent evaluation that may find results contrary to government policy, particularly in authoritarian or closed regimes. More open governments may, however, view evaluations and the dissemination of the findings as an important part of the democratic process.

Evaluations are also sensitive to political change. Three of the eight impact studies listed in table 4.2 were canceled because of political economy issues. Turnover in regimes or key posts within a counterpart government office and shifts in policy strategies can affect not only the evaluation effort, but more fundamentally the implementation of the program being evaluated. One example of this type of risk comes from the experience of a team working on the design and impact evaluation of a school-based management pilot in Peru as part of a World Bank financed primary education project. The team composed of Ministry of Education officials, World Bank staff, and international and local consultants had worked for over a year developing the school-based management models to be piloted, establishing an experimental design, designing survey instruments and achievement tests, and collecting baseline data on school characteristics and student achievement. Just prior to the pilot's introduction in the randomly selected schools, high level government officials canceled the school-based management experiment in a reaction to perceived political fallout from the pilot. A similar reform was introduced several years later, but without the benefit of a pilot test or an evaluation.

In Venezuela, an evaluation of a maternal and infant health and nutrition program was redesigned three times with three different client coun-

terparts as the government shifted responsibility for the evaluation from one agency to another. Each change was accompanied by a contract renegotiation with the private sector firm that had been identified to carry out the data collection and the majority of the analysis for the evaluation. When the legitimacy of the third government counterpart began to be questioned, the firm nullified the contract and the evaluation was abandoned. These incidents occurred during a period of political flux characterized by numerous cabinet reshufflings that ended with the collapse of the elected government serving as a counterpart for the project, so the evaluation was hardly alone in suffering from the repercussions of political instability. Nonetheless, in both the Peruvian and Venezuelan cases, it is sobering to reflect upon the amount of resources devoted to an effort that was never brought to fruition. A less dramatic example of the effect of political change on evaluation strategies comes from El Salvador, where the recognized success of a reform introduced in rural schools prompted the government to introduce a similar education reform in all of the urban schools at once, instead of randomly phasing in schools over time as originally planned. This decision eliminated the possibility of using an experimental design and left using a less-robust reflexive comparison as the only viable evaluation design option.

# Bibliography

The word *processed* describes informally reproduced works that may not be commonly available through libraries.


Abadie, A., J. Angrist, and G. Imbens. 1998. "Instrumental Variables Estimation of Quartile Treatment Effects." National Bureau of Economic Research Working Paper Series, No. 229:1–28.

Atkinson, Anthony. 1987. "On the Measurement of Poverty." *Econometrica* 55: 749–64.

Bamberger, Michael. 2000. *Integrating Quantitative and Qualitative Methods in Development Research.* Washington, D.C.: World Bank.

Barnes, Carolyn. 1966. "Assets and the Impact of Microenterprise Finance Programs." USAID AIMS Project Brief 6. Washington, D.C.: USAID.

Barnes, Carolyn, and Erica Keogh. 1999. "An Assessment of the Impact of Zambuko's Microenterprise Program in Zimbabwe: Baseline Findings." USAID AIMS Project Brief 23. Washington, D.C.: USAID.

Benus, Jacob, Neelima Grover, and Recep Varcin. 1998. *Turkey: Impact of Active Labor Programs.* Bethesda, Md.: Abt Associates.

Benus, Jacob, Neelima Grover, Jiri Berkovsky, and Jan Rehak. 1998. *Czech Republic: Impact of Active Labor Programs.* Bethesda, Md.: Abt Associates.

Besharov, Douglas J., Peter Germanis, and Peter H. Rossi. 1997. *Evaluating Welfare Reform: A Guide for Scholars and Practitioners.* College Park, Md.: University of Maryland.

Bindlish, V., R. Evenson, and Mathurin Gbetibouo. 1993. *Evaluation of T & V-Based Extension in Burkina Faso*. Washington, D.C.: World Bank.

Bloom, Howard S., Larry L. Orr, George Cave, Stephen H. Bell, Fred Doolittle, and Winston Lin. 1994. *The National JTPA Study: Overview, Impacts, Benefits, and Costs of Title II-A.* Bethesda, Md.: Abt Associates.

Blundell, Richard W., and R. J. Smith. 1993. "Simultaneous Microeconometric Models with Censoring or Qualitative Dependent Variables." In G. S. Maddala, C. R. Rao, and H. D. Vinod, eds., *Handbook of Statistics.* Vol. 11. Amsterdam: North Holland.

Bourguignon, Francois, Jaime de Melo, and Akiko Suwa. 1991. "Distributional Effects of Adjustment Policies: Simulations for Archetype Economies in Africa and Latin America." *World Bank Economic Review* 5 (2): 339–66.

Burtless, Gary. 1995. "The Case for Randomized Field Trials in Economic and Policy Research." *Journal of Economic Perspectives* 9 (Spring): 63–84.

Card, David, and Philip K. Robins. 1996. "Do Financial Incentives Encourage Welfare Recipients to Work? Evidence from a Randomized Evaluation of the Self-Sufficiency Project." National Bureau of Economic Research Paper 5701, August. Cambridge, Mass.: NBER.

Carvalho, Soniya, and Howard White. 1994. "Indicators for Monitoring Poverty Reduction." World Bank Discussion Papers 254. Washington, D.C.: World Bank.

Chen, Martha Alter, and Donald Snodgrass. 1999. "An Assessment of the Impact of Sewa Bank in India: Baseline Findings." USAID AIMS Project, Processed. Washington, D.C.: USAID.

Cohen, Monique, and Gary Gaile. 1997. "Highlights and Recommendations of the Virtual Meeting of the CGAP Working Group on Impact Assessment Methodologies." USAID AIMS Project. Washington, D.C.: Management Systems International.

———. 1998. "Highlights and Recommendations of the Second Virtual Meeting of the CGAP Working Group on Impact Assessment Methodologies." USAID AIMS Project. Washington, D.C.: Management Systems International.

Dar, Amit, and Indermit S. Gill. 1998. "Evaluating Retraining Programs in OECD Countries: Lessons Learned." *The World Bank Research Observer* 13 (February): 79–101.

Dar, Amit, and Zafiris Tzannatos. 1999. "Active Labor Market Programs: A Review of Evidence from Evaluations." World Bank Social Protection Discussion Paper 9901. Washington, D.C.: World Bank.

Dehejia, Rajeev H,. and Sadek Wahba. 1998. "Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs." NBER Working Paper Series 6586. Cambridge, Mass.: NBER http://www.nber.org/papers/w6586.

Dennis, Michael L., and Robert F. Boruch. 1989. "Randomized Experiments for Planning and Testing Projects in Developing Countries: Threshold Conditions." *Evaluation Review* 13 (June): 292–309.

Diagne, Aliou, and Manfred Zeller. 1998. "Determinants of Access to Credit and Its Impact on Income and Food Security in Malawi." Manuscript submitted to the International Food Policy Research Institute (IFPRI) Publication Review Committee for consideration as an IFPRI Research Report.

Diop, F., A. Yazbeck, and R. Bitran. 1995. "The Impact of Alternative Cost Recovery Schemes on Access and Equity in Niger." *Health Policy and Planning* 10 (3): 223–40.

Donecker, Jane, and Michael Green. 1998. *Impact Assessment in Multilateral Development Institutions.* London: Department for International Development.

Dunn, Elizabeth. 1999. "Microfinance Clients in Lima, Peru: Baseline Report for AIMS Core Impact Assessment." USAID AIMS Project, Processed. Washington, D.C.: USAID.

Edgecomb, Elaine L., and Carter Garber. 1998. "Practitioner-Led Impact Assessment: A Test in Honduras." USAID AIMS. Washington, D.C.: USAID.

Ezemenari, Kene, Anders Ruqvist, and K. Subbarao. 1999. "Impact Evaluation: A Note on Concepts and Methods." World Bank Poverty Reduction and Economic Management Network, Processed. Washington, D.C.: World Bank.

Foster, James, J. Greer, and Erik Thorbecke. 1984. "A Class of Decomposable Poverty Measures." *Econometrica* 52: 761–65.

Friedlander, Daniel, and Gayle Hamilton. 1993. *The Saturation Work Initiative Model in San Diego: A Five-Year Follow-Up Study.* New York: Manpower Demonstration Research Corporation.

Friedlander, Daniel, and Philip K. Robins. 1995. "Evaluating Program Evaluations: New Evidence on Commonly Used Nonexperimental Methods." *American Economic Review* 85 (September): 923–37.

Friedlander, Daniel, David H. Greenberg, and Philip K. Robins. 1997. "Evaluating Government Training Programs for the Economically Disadvantaged." *Journal of Economic Literature* 35 (December): 1809–55.

Fuller, Bruce, and Magdalena Rivarola. 1998. *Nicaragua's Experiment to Decentralize Schools: Views of Parents, Teachers and Directors.* Working Paper Series on Impact Evaluation of Education Reforms, paper no. 5. Washington, D.C.: World Bank.

Gaile, Gary L., and Jenifer Foster. 1996. "Review of Methodological Approaches to the Study of the Impact of MicroEnterprise Credit Programs." USAID AIMS Brief 2. Washington, D.C.: USAID

Glewwe, Paul, Michael Kremer, and Sylvie Moulin. 1998. "Textbooks and Test Scores: Evidence from a Prospective Evaluation in Kenya." Development Research Group, World Bank, October 28.

Goodman, Margaret, Samuel Morley, Gabriel Siri, and Elaine Zuckerman. 1997. *Social Investment Funds in Latin America: Past Performance and Future Role.* Washington D.C.: Inter American Development Bank. March.

Government of Denmark. 1995. *Methods for Evaluation of Poverty Oriented Aid Interventions.* Copenhagen: Ministry of Foreign Affairs.

Greene, W. H. 1997. *Econometric Analysis.* Hemel Hempstead, New Jersey: Prentice Hall Press.

Greenberg, David, and Mark Shroder. 1997. *The Digest of Social Experiments*, 2nd ed. Washington, D.C.: The Urban Institute Press.

Grosh, Margaret E., and Juan Muñoz. 1996. "A Manual for Planning and Implementing the Living Standards Measurement Study Survey." LSMS Working Paper #126. World Bank, Washington, D.C.

Grossman, Jean Baldwin. 1994. "Evaluating Social Policies: Principles and U.S. Experience." *The World Bank Research Observer* 9 (July): 159–80.

Habicht, J. P., C. G. Victoria, and J. P. Vaughan. 1999. "Evaluation Designs for Adequacy, Plausibility and Probability of Public Health Programme Performance and Impact." *International Journal of Epidemiology* 28: 10–18.

Harrell, Adele, Martha Burt, Harry Hatry, Shelli Rossman, Jeffrey Roth, and William Sabol. 1996. *Evaluation Strategies for Human Service Programs: A Guide for Policymakers and Providers.* Washington, D.C.: Urban Institute.

Heckman, James. 1997. "Instrumental Variables. A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations." *Journal of Human Resources* 32 (3): 441–61.

Heckman, James, R. Lalonde, and J. Smith. 1999. "The Economics and Econometrics of Active Labor Market Programs." In O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, vol. III. New York: Elsevier Science Publishing Co.

Heckman, James, and Richard Robb. 1985. "Alternative Methods of Evaluating the Impact of Interventions: An Overview." *Journal of Econometrics* 30: 239–67.

Heckman, James J., and Jeffrey A. Smith. 1995. "Assessing the Case for Social Experiments." *Journal of Economic Perspectives* 9 (2): 85–110.

Heckman, J., H. Ichimura, J. Smith, and P. Todd. 1998. "Characterizing Selection Bias Using Experimental Data." *Econometrica* 66: 1017–99.

Hicks, Norman. 1998. *Measuring the Poverty Impact of Projects in LAC.* World Bank Latin America and the Caribbean Region, Processed. Washington, D.C.: World Bank.

Holder, Harold D., Andrew J. Treno, Robert F. Saltz, and Joel W. Grube. 1997. "Summing Up: Recommendations and Experiences for Evaluation of Community-Level Prevention Programs." *Evaluation Review* 21 (April): 268–78.

Hollister, Robinson G., and Jennifer Hill. 1995. *Problems in the Evaluation of Community-Wide Initiatives.* New York: Russell Sage Foundation.

Hulme, David. 1997. "Impact Assessment Methodologies for Microfinance: Theory, Experience and Better Practice." Manchester,

U.K.: Institute for Development Policy and Management, University of Manchester.

International Food Policy Research Institute. 1998. *Programa Nacional de Educación, Salud, y Alimentación (PROGRESA): A Proposal for Evaluation* (with technical appendix). Washington, D.C.: IFPRI.

Jalan, Jyotsna, and Martin Ravallion. 1998a. "Income Gains from Workfare: Estimates for Argentina's TRABAJAR Program Using Matching Methods." Washington D.C.: Development Research Group, World Bank.

———. 1998b. "Transfer Benefits from Workfare: A Matching Estimate for Argentina." Washington, D.C.: Development Research Group, World Bank.

Jamison, Dean T., Barbara Serle, Klaus Galda, and Stephen P. Heyneman. 1981. "Improving Elementary Mathematics Education in Nicaragua: An Experimental Study of the Impact of Textbooks and Radio on Achievement." *Journal of Educational Psychology* 73 (4): 556–67.

Karoly, Lynn A., Peter W. Greenwood, Susan S. Everingham, Jill Houbé, M. Rebecca Kilburn, C. Peter Rydell, Matthew Sanders, and James Chiesa. 1998. *Investing in Our Children: What We Know and Don't Know about the Costs and Benefits of Early Childhood Interventions.* Santa Monica, Calif.: Rand.

Kemple, James J., Fred Doolittle, and John W. Wallace. 1993. *The National JTPA Study: Site Characteristics and Participation Patterns.* New York: Manpower Demonstration Research Corporation.

Khandker, Shahidur R. 1998. *Fighting Poverty with Microcredit: Experience in Bangladesh.* New York: Oxford University Press for the World Bank.

Killick, Tony. 1995. *IMF Programmes in Developing Countries, Design and Impact*. London: England.

King, Elizabeth, and Berk Ozler. 1998. *What's Decentralization Got to Do with Learning? The Case of Nicaragua's School Autonomy Reform.* Working Paper Series on Impact Evaluation of Education Reforms, paper no. 9. Washington, D.C.: World Bank.

King, Elizabeth, Berk Ozler, and Laura Rawlings. 1999. *Nicaragua's School Autonomy Reform: Fact or Fiction?* Washington, D.C.: World Bank.

Kish, Leslie. 1965. *Survey Sampling.* New York: John Wiley and Sons.

Levinson, James F., Beatrice Lorge Rogers, Kristin M. Hicks, Thomas Schaetzel, Lisa Troy, and Collette Young. 1998. *Monitoring and Evaluation: A Guidebook for Nutrition Project Managers in Developing Countries.* Medford, Mass.: International Food and Nutrition Center, Tufts University School of Nutrition Science and Policy, March.

Manski, Charles, and Irwin Garfinkel, eds. 1992. *Evaluating Welfare and Training Programs.* Cambridge, Mass.: Harvard University Press.

Manski, Charles, and Steven Lerman. 1977. "The Estimation of Choice Probabilities from Choice-Based Samples." *Econometrica* 45: 1977–88.

McKay, H., A. McKay, L. Siniestra, H. Gomez, and P. Lloreda. 1978. "Improving Cognitive Ability in Chronically Deprived Children" *Science 2000*(21): 270–78.

Meyer, Bruce D. 1995. "Natural and Quasi-Experiments in Economics." *Journal of Business and Economic Statistics*, 13 (April): 151–161.

Miles, Matthew B., and A. Michael Huberman. 1994. *Qualitative Data Analysis.* London: Sage Publications.

MkNelly, Barbara, and Karen Lippold. 1998. "Practitioner-Led Impact Assessment: A Test in Mali." USAID AIMS Brief 21. Washington, D.C.: USAID

MkNelly, Barbara, and Christopher Dunford (in collaboration with the Program in International Nutrition, University of California, Davis). 1998. "Impact of Credit with Education on Mothers' and Their Young Children's Nutrition: Lower Pra Rural Bank Credit with Education Program in Ghana." Freedom from Hunger Research Paper No. 4, March.

Moffitt, Robert. 1991. "Program Evaluation with Nonexperimental Data." *Evaluation Review* 15 (3): 291–314.

Mohr, Lawrence B. 1995. *Impact Analysis for Program Evaluation*, 2nd ed. Thousand Oaks, Calif.: Sage Publications.

———. 1999. "The Qualitative Method of Impact Analysis." Processed.

Morduch, Jonathan. 1998a. "The Microfinance Schism." Harvard Institute for International Development, Development Discussion Paper 626, February.

———. 1998b. "Does Microfinance Really Help the Poor? New Evidence from Flagship Programs in Bangladesh." Harvard University Institute for International Development, Processed, June.

———. 1999. "The Microfinance Promise." *Journal of Economic Literature* 37(4): 1569–1614.

Newman, John, Menno Pradhan, Laura Rawlings, Ramiro Coa, and Jose Luis Evia. 2000. "An Impact Evaluation of Education, Health, and Water Supply Investments of the Bolivia Social Investment Fund." Processed.

Newman, John, Laura Rawlings, and Paul Gertler. 1994. "Using Randomized Control Designs in Evaluating Social Sector Programs in Developing Countries." *The World Bank Research Observer* 9 (July): 181–202.

Operations Evaluation Department, World Bank. 1994. "An Overview of Monitoring and Evaluation in the World Bank." June 30.

Poppele, J. S. Summarto, and L. Pritchett. 1999. "Social Impacts of the Indonesia Crisis: New Data and Policy Implications." Social Monitoring Early Response Unit, World Bank, Washington, D.C. Processed.

Pradhan, Menno, Laura Rawlings, and Geert Ridder. 1998. "The Bolivian Social Investment Fund: An Analysis of Baseline Data for Impact Evaluation." *World Bank Economic Review* 12 (3): 457–82.

Ravallion, Martin. 1994. *Poverty Comparisons, Fundamentals in Pure and Applied Economics*, vol. 56. Harwood Academic Publishers.

————-. 1999. *Monitoring Targeting Performance When Decentralized Allocations to the Poor Are Unobserved*. World Bank, Washington, D.C. Processed.

Ravallion, Martin, and Quentin Wodon. 1998. *Evaluating a Targeted Social Program When Placement Is Decentralized.* World Bank Policy Research Working Paper 1945. Washington, D.C.: World Bank.

Ravallion, Martin, Dominique van de Walle, and Madhur Gautam. 1995. "Testing a Social Safety Net." *Journal of Public Economics* 57: 175–99.

Rawlings, Laura. 2000. "Assessing Educational Management and Quality in Nicaragua." In Bamberger: *Integrating Quantitative and Qualitative Methods in Development Research*. Washington, D.C.: World Bank.

Rebien, Claus C. 1997. "Development Assistance Evaluation and the Foundations of Program Evaluation." *Evaluation Review* 21 (August): 438–60.

Revenga, Ana, Michelle Riboud, and Hong Tan. 1994. "The Impact of Mexico's Retraining Program on Employment and Wages." *World Bank Economic Review* 8 (2): 247–77.

Rivers, Douglas, and Quang H. Vuong. 1988. "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models." *Journal of Econometrics* 39: 347–66.

Robinson, Sherman. 1991. "Macroeconomics, Financial Variables, and Computable General Equilibrium Models." *World Development*, 19(11): 1509–25.

Rosenbaum, P., and D. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70: 41–55.

————-. 1985. "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score." *American Statistician* 39: 35–39.

Rossman, Gretchen B., and Bruce L. Wilson. 1985. "Numbers and Words: Combining Quantitative and Qualitative Methods in a Single Large-Scale Evaluation Study. *Evaluation Review* 9 (3): 627–43.

Sahn, David, Paul Dorosh, and Stephen Younger. 1996. "Exchange Rate, Fiscal and Agricultural Policies in Africa: Does Adjustment Hurt the Poor?" *World Development*, 24(4): 719–47.

Sebstad, Jennifer, and Gregory Chen. 1996. "Overview of Studies on the Impact of MicroEnterprise Credit." USAID AIMS, June. Washington, D.C.: USAID

Sharma, Manohar, and Gertrud Schrieder. 1998. "Impact of Finance on Food Security and Poverty Alleviation—A Review and Synthesis of Empirical Evidence." Paper presented to Workshop on Innovations in Micro-Finance for the Rural Poor: Exchange of Knowledge and Implications for Policy, Ghana, November.

Smith, William. 1998. *Group Based Assessment of Change: Method and Results.* Hanoi, Vietnam: ActionAid.

Subbarao, K., Kene Ezemenari, John Randa, and Gloria Rubio. 1999. *Impact Evaluation in FY98 Bank Projects: A Review.* World Bank Poverty Reduction and Economic Management Network, Processed, January.

Tan, J.P, J. Lane, and G. Lassibille. 1999. "Schooling Outcomes in Philippine Elementary Schools: Evaluation of the Impact of Four Experiments." *World Bank Economic Review*, September.

Taschereau, Suzanne. 1998. *Evaluating the Impact of Training and Institutional Development Programs, a Collaborative Approach.* Economic Development Institute of the World Bank, January.

USAID. 1984. "Study of Low Cost Housing." New Delhi: USAID. June.

———. 1995. "Assessing the Impacts of MicroEnterprise Interventions: A Framework for Analysis." MicroEnterprise Development Brief, June.

Valadez, J., and M. Bamberger, ed. 1994. "Monitoring and Evaluating Social Programs in Developing Countries." Economic Development Institute of the World Bank Series, World Bank, Washington, D.C.

van de Walle, Dominique. 1999. "Assessing the Poverty Impact of Rural Road Projects." World Bank. Processed.

Walker, Ian, Rafael del Cid, Fidel Ordoñez, and Florencia Rodriguez. 1999. "Evaluación Ex-Post del Fondo Hondureño de Inversión Social (FHIS 2)." ESA Consultores, Tegucigalpa, Honduras.

Weiss, Carol H. 1998. *Evaluation.* Upper Saddle River, New Jersey: Prentice Hall.

Wilson, Sandra Jo. 1998. "A Multi-Method Proposal to Evaluate the Impact of CGIAR-Instigated Activities on Poverty Reduction." Paper presented to the Annual Conference of the American Evaluation Association, November.

World Bank, Poverty and Human Resources Division, Policy Research Department. 1994a. *Impact Evaluation of Education Projects Involving Decentralization and Privatization.* Working Paper. Washington, D.C.: World Bank.

———, Learning and Leadership Center and Operations Policy Department. 1996. *Handbook on Economic Analysis of Investment Operations.* Washington, D.C.: World Bank.

———, Operations and Evaluation Department. 1998a. *1998 Annual Review of Development Effectiveness.* Washington, D.C.: World Bank.

———, Poverty Reduction and Economic Management, Latin America and Caribbean Region. 1998b. *Measuring the Poverty Impact of Projects in LAC.* Washington, D.C.: World Bank.

———, Operations and Evaluation Department. 1999. *World Bank Agricultural Extension Projects in Kenya: An Impact Evaluation.* Report No. 19523. Washington, D.C.: World Bank.

_____, Poverty Reduction and Economic Management Unit, 1999. *Strategy for Impact Evaluation.* Washington, D.C.: World Bank.

Wouters, A. 1995. "Improving Quality through Cost Recovery in Niger." *Health Policy and Planning* 10 (3): 257–70.

Zeller, Manfred, Akhter Ahmed, Suresh Babu, Sumiter Broca, Aliou Diagne, and Manohar Sharma. 1996. "Security of the Poor: Methodologies for a Multicountry Research Project." IFPRI Discussion Paper 11.