



International Experience in Impact Evaluation

Impact Evaluation Workshop
Arianna Legovini, World Bank
Mombasa, Aug 31, 2005

Objective of this presentation is to

- Share experiences on impact evaluation that:
 - Exemplify the value of impact evaluation for improving public policy
 - Provide guidance for the project-specific discussions during this workshop

Impact evaluation *is*

- Measuring the effect of a development activity on a beneficiary population controlling for all other factors that might have affected that population during the evaluation period.

Impact *is*

- The portion of the change in any outcome (short, medium or long term) that can be attributed to that development activity.

What can Impact Evaluation do for us? Improve use of public finances

- Create knowledge of what works:
 - Measure cost-effectiveness
 - Rank policy alternatives
 - Identify problems in projects and government programs
- Apply that knowledge:
 - ➔ Perfect projects over time by doing more of what works and less of what doesn't (project management tool)
 - ➔ Shift fiscal allocations over time towards cost-effective policies and away from ineffective policies (macro policy tool) will increase overall effectiveness in the use of fiscal resources

Measure cost effectiveness: health & education in Kenya

- Based on randomized experiments in Busia schools, Kremer (2003) estimates the cost-effectiveness of different instruments in delivering one extra year of schooling per student:
 - Provision of school uniforms costs \$99 per year per child
 - School feeding costs \$36 per year per child
 - Deworming treatment a mere \$3.50 per year per child
- In other words, \$10000 of public resources put in:
 - Deworming will keep 2800 additional children in school for an additional year
 - Uniforms or school feeding will keep only 100-277 children in school for an additional year
- These results have been incorporated in the Kenya education sector strategy and Education SWAP

Evaluate program before expanding it: Mexico

- Mexico launched *PROGRESA* in 1997, a Conditional Cash Transfer program—mothers receive transfers conditional on investment in human capital of children (education, health, nutrition).
- In 1998 randomly assigned 506 villages to treatment and control. Delayed by two years program in control villages.
- Results in treated communities after two years:
 - Share of households living on poverty declined by 8% in treated communities, poverty gap by 30%, and poverty depth by 45%;
 - One third of the children not previously enrolled attended school (90% or more of the time);
 - Children aged 0-5 experienced a 12% decrease in incidence of disease; and
 - Annual mean growth of children aged 12-36 months increased by 1cm (or 16 % more).
- *PROGRESA was expanded country wide.*
- **It survived a change in administration (to the opposition)**

Is any Impact Evaluation done in Africa?

- Growing body of literature (see Legovini 2004)
- Most studies in agriculture and rural development (40%) and human development (50%) sectors
- More efforts needed in infrastructure, financial sector, public services and private sectors

International experience: relevant examples for participants

1. Community Driven
Development/Social Funds
2. Urban infrastructure/Slum
upgrading
3. Agricultural services/extension
4. Private and financial sector
5. Public services

1. Evaluating the impact of a Community Driven Development (CDD) project

Evaluating the impact of a CDD project

- CDDs empower communities into selecting developmental alternatives
- Difficulties for evaluating impact of CDD include:
 - Communities self-selected into participation
⇒ Selection bias and treatment group unknown beforehand ⇒ sample design issues
 - Interventions are community-specific ⇒ outputs and outcomes are undetermined ex ante ⇒ instrument design issues, implications for sample size

Evaluating the impact of a CDD project

- Qualitative methods (to investigate empowerment and social capital issues) +
- Sampling, Experimental
 - Randomly assigns areas eligible to participate in the program [requires agreement with client and baseline surveys] e.g. Bolivia (Newman et al. 2002)
- Sampling, Non-experimental
 - Assign areas that will not be eligible for treatment
 - Use baseline to “match” observations
- Instrument design: include all likely outputs (schools, health units, wells and boreholes, etc) and good range of outcomes
- Surveying options: CDD-specific baseline and follow up surveys or use existing planned ho. surveys and over-sample areas of interest (timing??)

CDD: An example from Bolivia's Social Investment Fund

- **PROJECT**

- SIF created in 1991 to improve coverage and quality of basic services in education, health, water and sanitation.

- **REFERENCE**

- Newman, John, Meno Pradhan, Laura B. Rawlings, Geert Ridder, Ramiro Coa, and Jose Luis Evia. 2002. "An Impact Evaluation of Education, Health and Water Supply Investments by the Bolivian Social Investment Fund." *The World Bank Economic Review* 16(2): 241-274

CDD: An example from Bolivia's SIF

- **DESIGN** The evaluation uses different methods for different interventions in two regions, the Chaco region and the Resto Rural.
 - **Education:**
 - In Chaco region randomization was done using a school quality index. Worst off communities were automatically designated as eligible, and better off automatically designated as ineligible. “Middle” communities were included in the randomization (200 schools: 114 control, 86 treatment).
 - In Resto Rural, comparison group of non-SIF schools was constructed using a two-step matching process based on observable characteristics of communities (from a recent census) and schools (from administrative data). Propensity score matching on 1998 data.
 - The analysis used difference-in difference impact evaluators.

CDD: An example from Bolivia's SIF

- DESIGN

- Health:

- Initially impact was going to be measured via reflexive comparison, i.e. comparing values before and after the intervention for the same population.
 - OPPORTUNITY Financial constraints prevented SIF from investing in all health centers in the Resto Rural, thus creating a control group that had the same initial mean characteristics as the treatment group but which did not receive the project. A control group was identified using propensity score matching.
 - This allowed difference-in-difference estimators instead of difference only estimators.

- Water:

- The evaluation constructed the comparison group from the health sub-sample using simple matching to identify similar non-beneficiaries.

CDD: An example from Bolivia's SIF

- DESIGN cont.
 - The data for the evaluation were collected through a baseline survey in 1993 and a follow-up data survey in 1997-1998. Both surveys collected data from 5 provinces in the Chaco region and 17 provinces Resto Rural. Five types of data collection were used: household surveys, facilities surveys, community surveys, water quality samples, and student achievement tests.

CDD: An example from Bolivia's SIF

SELECTED RESULTS

- Education
 - Significant increase in the fraction of schools with sanitation facilities, number of textbooks per student, and student attendance
 - Significant decrease in the number of students per classroom and per teacher, and school dropout rate.
- Health
 - Significant improvements in health clinic characteristics: number of beds, sanitation facilities, and patient rooms, and availability of medical supplies.
 - Significant increase in the proportion of women receiving prenatal care and the proportion of cough cases treated.
 - Significant decrease in child mortality and under-age-five mortality.

2. Evaluating the impact of urban infrastructure

Evaluating the impact of urban infrastructure

- The objective of slum upgrading programs is to improve living conditions in low-income settlements by upgrading basic infrastructure, e.g., roads and pathways, water supply, drainage, sanitation, and street lighting.
- Issues for impact evaluation include:
 - disentangling the effects of these different interventions (implications for sample size, instrument design)
 - tracking the effect of increased real estate/neighborhood values on the welfare of owners and renters—the latter may be pushed out (of the areas and of the sample) and end up worse off, results may be biased upwards

Urban infrastructure: an example from Tanzania

Reference: Terms of Reference for the Design and Implementation of an Impact Evaluation of the CIUP (LGSP), Arianna Legovini (2005)

Project :

- The Community Infrastructure Upgrading Program of LGSP will provide a community determined package for upgrading basic infrastructure and services in 16 selected low-income settlements of Dar es Salam.
- DAWASA will provide water upgrades in the 16 CIUP communities and 15 additional communities.

Design:

- Representative sample of households from
 - 16 CIUP communities,
 - 15 DAWASA only communities, and
 - ___ non-beneficiary communities.

Urban infrastructure: an example from Tanzania

- Observations will be matched using baseline survey (2005)
- Households leaving the areas will be tracked over time and kept in the sample (using monetary incentives).
- Follow up survey (2007/08)
- Impact will be measured using difference-in-difference estimators across a wide range of welfare indicators.
 - CIUP vs non-beneficiary will provide estimates of impact for whole package.
 - CIUP vs DAWASA, for the package excluding water.
 - DAWASA vs non-beneficiary, for water only.

3. Evaluating the impact of agricultural services

Evaluating the impact of agricultural services

- Agricultural services are extension, marketing and credit services provided to farmers to e.g. strengthen technology adoption, use of fertilizer, diversification of crops, adoption of marketing strategies, etc.
- Difficulties in evaluating impact include self-selection of participants, heterogeneity in the quality of service, poor market signals when service prices are set at zero, and biased data when collection administered by service providers

Agricultural services: an example from Kenya

Reference: Finding Missing Markets: An Agricultural Brokerage Intervention in Kenya, Nava Ashraf, Xavier Gine (World Bank), Dean Karlan.

Design

- The objective of the study is to test the effectiveness and sustainability of DrumNet, an agricultural service program with credit.
- Randomized design to test separately the impact of credit from the impact of agricultural extension and marketing services.
- Hypotheses: Technical assistance will encourage faster adoption of high-return crops, resulting in higher yield, sales and income levels. Credit recipients will have higher profits, hence indicating credit constraints as an obstacle to growth.

Agricultural services: an example from Kenya

Design

- Field visit produced the list of all 96 horticulture self-help groups (SHGs) in Gichugu registered since 2000—a total of approx. 3,000 farmers.
- Identification of 20-40 SHGs with combined membership of 750 individuals (20-40 members per group)
- Random assignment of the SHGs into three experimental groups of 250 participants each:
 - 1) control,
 - 2) all DN services, and
 - 3) DN services except credit.
- Randomization to ensure that the three groups look alike ex-ante along several key variables (landholdings, farming experience, crop mix, access to credit and infrastructure)
- Comparing outcomes between the two treatment groups to measure the effect of credit.

Agricultural services: an example from Kenya

Design

- Official start of the experiment March, 2004.
- Baseline survey administered to all 750 members of the selected SHGs.
- Follow-up survey expected 2006.
- Difference-in difference to assess the broader economic impact of a particular treatment.
- Primary costing analysis to examine the accounting and cash flows of the DrumNet enterprise and quantify the subsidy, if any, required on an ongoing basis.
- Activity-based costing exercise to understand cost drivers and hence expansion obstacles and optimal pricing model.

4. Evaluating the impact of private and financial sector initiatives

Evaluating the impact of private and financial sector initiatives

- **The objective of PS initiatives may include improving investment climate, reducing cost of doing business, or relaxing constraints to growth of SMEs through provision of business services and credit**
- **Difficulties in evaluating impact include attributing change to institutional changes that affect all businesses at the same time (e.g. one stop investment), or self-selection in the case of business services or credit**

Private Sector: an experiment to encourage registration of informal firms' in São Paulo, Brazil

Motivation:

- Many firms in Brazil remain informal. Registering a company is a difficult process in Sao Paulo which takes as long as 152 days. Numerous policies could be used to encourage formality. What effects would these policies have?

Reference: Marianne Bertrand (U. of Chicago), Simeon Djankov (WB), and Sendhil Mullainathan (Harvard U.)

Experiment:

- The objective of the experiment is to encourage business registration

Private Sector: an experiment in Sao Paulo, Brazil

The study proceeds in three steps.

- Questionnaire is delivered to about 1,000 businesses in São Paulo.
- Business owners willing to become formal are invited to come to a session. Participants are randomly assigned to:
 - *Control group*: Participants in this group are given a talk by a prominent local business man.
 - *Treatment group*: Participants in this group receive the above plus a variety of treatments:
 - Encouragement to become formal, including testimonials on the benefits of incorporation.
 - Information and help on the process and forms needed for registration.
 - Provision of monetary resources to help participants pay for registration expenses.
 - **Reminders and follow-ups.**

4

Private Sector: an experiment in Sao Paulo, Brazil

- A follow up survey about six months after the seminar sessions is carried out to evaluate the impact of the treatment on registration and on economic outcomes.

Financial Sector: an experiment to test loan uptake in RSA

Reference: Marketing Effects in a Consumer Credit Market, Marianne Bertrand, Dean Karlan, Sendhil Mullainathan, Eldar Shafir, Jonathan Zinman

Experiment:

- The study investigates marketing effects on loan acceptance.
- Researchers send out letters to South African bank customers.
- Various marketing factors (such as the acceptance deadline date, the photograph of the bank manager, and the language describing acceptable use of money) are randomly varied.
- Results of this project to be compared against the interest rate effects on loan acceptance.
- Researchers will focus on the way that such psychological factors can influence financial decisions.

5. Evaluating the impact of using Public Expenditure Tracking Surveys and citizens report cards for improving service delivery

Evaluating the impact of using PETS and report cards for improving service delivery

- Many countries and CSOs have used citizen report cards and public expenditure tracking surveys to improve service delivery by empowering service users and making service providers more accountable.
- The results have included increases in the proportion of fiscal resources reaching facilities (e.g. Uganda education) and in satisfaction with services (e.g. Bangalore, India). These are attributed to information dissemination to users and providers of services.

5

Tracking public expenditure: an example from Uganda's primary education

- Starting in early 1990s, Public Expenditure Tracking Surveys (PETS) in primary education analyzed flows of funds through tiers of government
- **In 1991-95 only 13% of earmarked funds reached schools**
- Government began publishing monthly inter-governmental transfers in newspapers, making radio announcements, and requiring schools to post information on their walls.
- **1999-2000 PETS showed an increase of funds reaching schools to 80-90%.**

5

Monitoring service delivery with report cards: an example from Bangalore, India

- In 1993, the Public Affairs Center started to collect feedback from users on their perceptions of quality and efficiency of public services : the 'report card' rated performance of all major service providers in the city.
 - 10.5 % of households were satisfied with services and 37.5 % of households were dissatisfied
- This exercise was repeated in 1999, and replicated in many other Indian cities and states.
 - Satisfaction increased to 40.1 % of households and dissatisfaction fell to 17.9 % of households
- Reference: Participatory Approaches in Budgeting and Public Expenditure Management, World Bank

The next step--Evaluating the impact of using report cards for improving health service delivery in Uganda

Reference: Impact evaluation of citizen report card at the community level, Ritva Reinikka

Design

- **Objective:** evaluate the impact of the citizen report card at the community level (CRCCL) on service delivery performance and outcomes using an experimental design. The source of identification will thus come directly from a randomized experiment.
- The study randomly assigns government health clinics (say, 20-25) and the communities they are serving to participate in the CRCCL to the treatment group, and provided with advocacy and awareness training.

The next step--Evaluating the impact of using report cards for improving health service delivery in Uganda

Design cont.:

- Another group (again, say, 20-25 clinics) is assigned to the control group.
- The treatment effect is derived by comparing outcomes (in different dimensions) between the treatment and control groups.
- It is important to ensure that, ex ante, the treatment and control groups are similar. The provider and user surveys will be implemented before and after the intervention (i.e., interface process).

Impact evaluation and Bank projects: first best

- Impact evaluation is incorporated in the design of Bank operation
- Sufficient funding is available for it during both project preparation and implementation phases
- Preparation:
 - Impact evaluation team collaborates with project team in project design, incl. selection of beneficiaries (random assignment of treatment and control)
 - Impact evaluation team designs analytical framework and survey instruments
- Implementation:
 - Baseline data collection in the field at or before project effectiveness
 - Follow-up data collection at some frequency thereafter
 - Data analysis, dissemination and policy feedback

Impact evaluation and Bank projects: first best

- Impact evaluation results are discussed with project teams, country teams and sectors, and incorporated in CAS
- Impact evaluation results are disseminated to government and incorporated in policy discussions and program design



THANK YOU