

International Workshop on Impact Evaluation for Development

DAC-OECD, Paris

# Some Issues and Questions Identified in JBIC's Impact Evaluation Studies

November 15, 2006

Seiro Ito<sup>†</sup>

Japan Bank for International Cooperation  
and Institute of Developing Economies

---

<sup>†</sup> ✉ [itohse@ide.go.jp](mailto:itohse@ide.go.jp)

# 1 JBIC's Recent Impact Evaluation Studies 1

---

## 1 JBIC's Recent Impact Evaluation Studies

**Bangladesh:** impact of Jamuna Multipurpose Bridge Project on poverty reduction.

**India:** role of infrastructure in poverty reduction.

**Peru:** improvement of living environment and livelihood in poor communities.

- In Peru, we surveyed 2400 households in Sierra and in Serva.
- And we were wondering if the answers were correct...

Example quotes from the Peruvian questionnaire:

- In the last 2 months, did [NAME] have any of the following illnesses?

Diarrhea / Respiratory infection /  
Parasitic disease / Other /  None .

- ➔ There may be a *perception error* that they do not correctly understand the symptoms of these illnesses (see Strauss and Thomas, 1998 for a discussion about self-reported morbidity).

Example quotes from the Peruvian questionnaire:

● Is the water potable?  Yes / No / Don't know.

● What is your opinion on the water quality for drinking?

Good /  OK / Bad.

☞ Nobody chooses 'bad' given that they are drinking.

➔ There may be a *self-comforting bias* toward the better quality.

Potable water. No diarrhea.

No respiratory problems?

No parasitic problems?

Example quotes from the Peruvian questionnaire:

- How many acres is the plot No. 1?

\_\_\_\_\_ acres / Other (2 lines).

- How much [FRUIT NAME] did you produce in the last season?

\_\_\_\_\_ KGs / Other (whatever necessary).

- ➔ There may be *different units of measurement*.
- ➔ There can be several *different modes of production*, such as harvest without planting nor sales.



Take as much as you want.

## 2 Need of Scientific/Objective Measurement 9

---

### 2 Need of Scientific/Objective Measurement

- ❑ Some of the self-reported variables are error-ridden.

Q1. When should we use the scientific/objective measurement?

- ① Size of errors is considerable (errors-in-variables problem).

- ✎ Irrigation water use, food intake and physical workload (blood tests).

- ② The poor is more likely to provide inaccurate answers.

- ✎ Birth dates, school attendance.

## 2 Need of Scientific/Objective Measurement **10**

---

Q1. When should we use the scientific/objective measurement? (continued)

③ Content is difficult to perceive or to know.

✍ Respiratory diseases, parasitic diseases, water quality, air quality.

④ Social norms prevent respondents from reporting truthfully.

✍ Gynecological diseases, psychiatric conditions, HIV/AIDS, other sexually transmitted diseases (STDs).

## 2 Need of Scientific/Objective Measurement **11**

---

But there are problems with the scientific measurement:

- High costs, even only for the testing.
- Low logistical feasibility (or high logistic costs).
  - ① Cold chains: blood and water samples must be stored at below certain temperature.
  - ② Time limits: specimen must be delivered within certain hours.
- ▣➔ Most remote areas, where measurement is most needed, are least likely to be covered.
- ▣➔ Must consider the costs and benefits of scientific measurement.

## 2 Need of Scientific/Objective Measurement **12**

---

A cold chain through there...

## 2 Need of Scientific/Objective Measurement **13**

---

Additional caveats for the scientific measurement of health-related variables:

- ✎ Can only measure the chronic health conditions.
- ✎ Ability to perform the activities of daily lives (ADLs) are better for the acute symptoms.

□ But some of the errors can be avoided without scientific measurement if we know:

- local measurement units.
- local modes of production.
- local perception (errors).
- other self-reporting biases.

E.g., an air quality question can be changed to exposure to pollutant, and can be substituted with some of the following questions that are suitably modified to get the accurate information:

- How often do you use wood/charcoal for cooking?
- Who lights up the fire?
- Who prepares the food?
- Who are with [THE PREPARER] during cooking?

## 2 Need of Scientific/Objective Measurement **15**

---

➔ Knowing the *local context* is crucial in writing up the questionnaire and designing the survey.

Q2. How do we collect these error-prone data?

- With the local partners.

Q3. Who should be the impact evaluator?  
Can we be one?



## 3 Local Context and Impact Evaluator Choice **16**

---

### 3 Local Context and Impact Evaluator Choice

- ❑ JBIC has been trying to localize impact evaluation.
- ❑ Difficulty: partner government's limited capacity.

So Q3 becomes, if we are to reflect real world problems:

Q3. Who should be the impact evaluator?

- To best achieve the goals.
- Under:
  - technological constraints, and,
  - resource constraints.

4 A Decision Framework

In seeking answers to it, we must ask:

Q3-1. What are the **impact evaluation goals**?

Q3-2. What is the **impact evaluation technology**?

Q3-3. What are the existing **resources**?

Q3-4. What are the **constraints**?

[AN IMPACT EVALUATOR'S PROBLEM (in words)]

Maximize the achievement of impact evaluation goals by optimally choosing the resources used for impact evaluation technology, subject to various constraints.

### 5 Impact Evaluation Technology

- Inputs:

1. Technical capacity ( $T$ ).
  2. Local context ( $L$ ).
  3. Past data ( $D$ ) as baseline or a reference.
  4. Measurement devices ( $M$ ).
- ✎  $T, L$  are combined at the impact evaluator level.

- Outputs:

1. Treatment effect estimation.
2. Theory-based interpretations.

### 6 Existing Resources

#### 1. Human and network resources (impact evaluators):

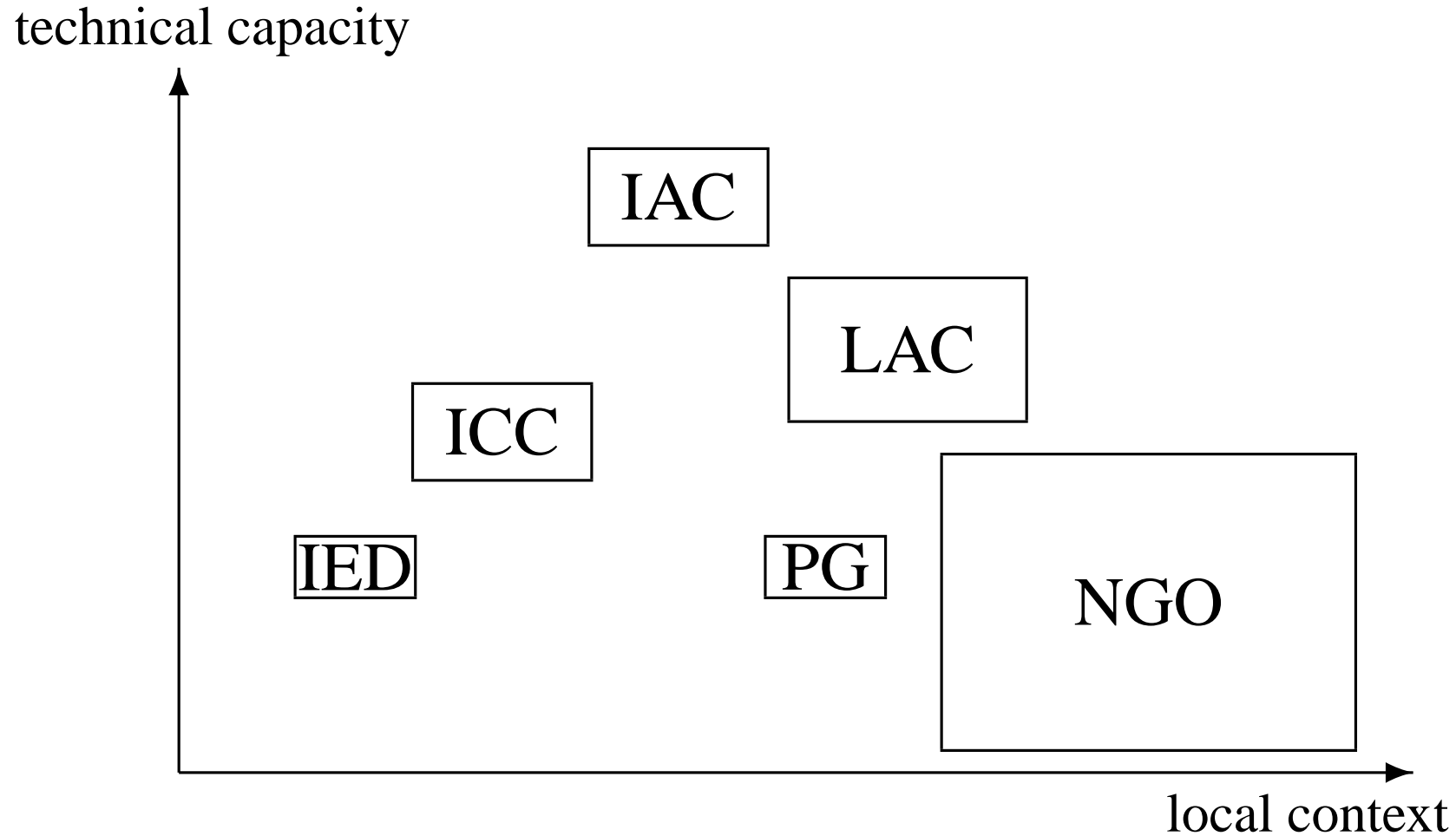
- Partner government (PG).
- Local academic consultants (LAC).
- NGOs.
- International commercial consultants (ICC).
- International academic consultants (IAC).
- Internal evaluation department (IED).

#### 2. Material resources:

- Past (related) evaluations.
- Past data.
- Measurement devices.

➔ Need to optimize the mix of resources.

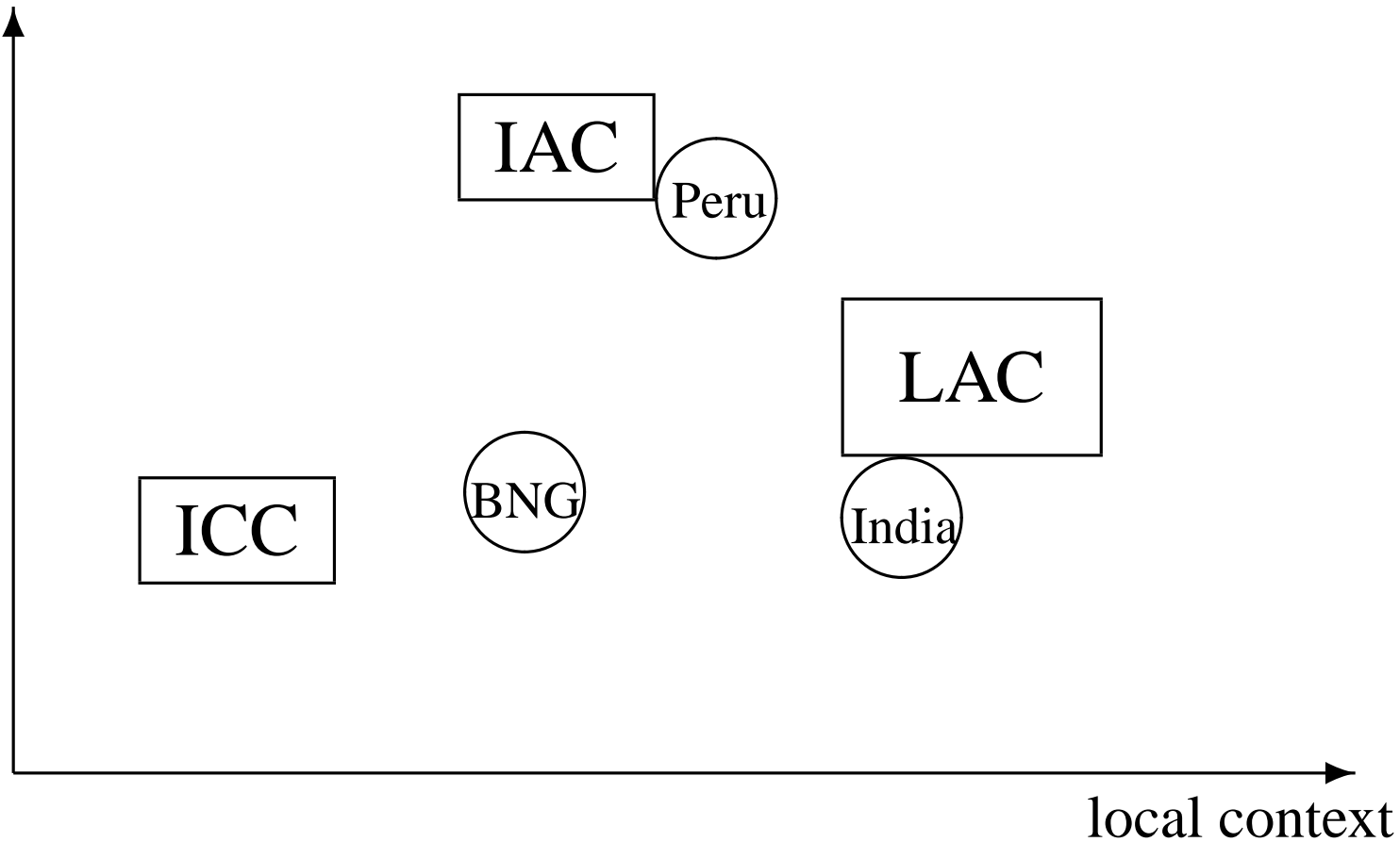
## COMPARATIVE ADVANTAGE OF IMPACT EVALUATORS (SPECULATIVE)



size of  = cost-efficiency

## COMBINED COMPARATIVE ADVANTAGE OF JBIC STUDIES (SPECULATIVE)

technical capacity



### 7 Constraints

#### 1. Budget constraint.

We will discuss this later.

## 2. Material constraints:

Q4. How can insufficient baseline data be coped with *ex post*?

① Existing information.

▣▣▣▣➤ WHERE IS IT?

② Recall information.

▣▣▣▣➤ ERROR-RIDDEN.

▣▣▣▣➤ MARTIN RAVALLION'S EXPERIMENT.



Q4. How can insufficient baseline data be coped with *ex post*? (continued)

③ Propensity score matching, provided that there are no omitted variables.

▣▣▣▣➤ UNLIKELY TO HOLD.

④ Regression discontinuity design, provided that the policy rules are strictly implemented.

▣▣▣▣➤ OTHER UNOBSERVABLE FACTORS MAY DOMINATE. ▣▣▣▣➤ CASES ARE LIMITED.

⑤ Pipeline matching, provided that the implementation order is not selective.

▣▣▣▣➤ UNLIKELY, IF NOT DELIBERATELY RANDOMIZED.

Q4. How can insufficient baseline data be coped with *ex post*? (continued)

⑤ Extrapolation, provided that all the regressors (for smoothing and treatment effect estimation) are available for the study area.

▮▮▮▮➔ NEED TO COLLECT MORE VARIABLES.

➔ Cannot be coped. Need more data or assumptions are not plausible. Collect the baseline. *Just do it.*

➔ Why are we evaluating *ex post* to begin with? *The donor must work with the partner government in collecting the baseline* to come up with a candidate project list.

Q5. How do we prepare the baseline in advance under the project implementation cycle?

① Universal baseline: all projects collect the simple baseline of time-varying information.

✓ No selectivity in evaluation. An aggregated impact.

✗ Costly, prolongs project cycle, less rigorous.

② Selective baseline: stratified random selection of projects for extensive data collection.

✓ More rigorous evaluation.

✗ An effort bias due to announcement prior to implementation. Selectivity due to not covering all strata.

⇒ An example of selectivity: I was not randomly assigned to study the Peruvian case.

Q6. Which projects should be evaluated?

To be more careful about selectivity, one may have to ask:

Q7. Which projects have been implemented?  
What determines the project implementation?

Q8. Which projects should be implemented?  
What should determine the project implementation?

➔ Local political economy dictates to a considerable extent.

To get some idea, we can:

① Estimate the political-economy models.

▮▮▮▮➔ Needs the knowledge of local context.

② Estimate a ‘causal’ poverty map:  $\text{Poverty} = f(\text{local policies, other exogenous variables})$ .

▮▮▮▮➔ Needs the knowledge of local context.

▮▮▮▮➔ For policymakers: a guidance on priority.

▮▮▮▮➔ For researchers: a placement rule.

### 8 Current Achievements

Q9. What can we know in the current framework?

We must address the selectivity on what population we are drawing conclusion.

The project we evaluate is a project:

- purposefully *chosen by the impact evaluator,*
- out of all the projects purposefully *chosen by the policymakers,*
- out of all the possible projects.

For that project, we can get:

- ① Distribution of final outcomes (total derivatives).
  - ▣▣▣▣ Treatment effect estimation.
- ② Bottlenecks and catalysis (partial derivatives).
  - ▣▣▣▣ Finer factorization of final outcomes.
  - ▣▣▣▣ Requires finer and further uses of experiments/instruments, which calls for extensive knowledge of local context.
  - ▣▣▣▣ Can be done only to a very limited extent.
  - ▣▣▣▣ Theory-based method can also serve this purpose.
  - ▣▣▣▣ There is no formal way of mixing the two.

After obtaining the results, we need to interpret why we had them.

- ▣➔ Need the knowledge of local context in interpreting the results.

So, the knowledge of local context is necessary in:

- ① Designing a survey.
- ② Writing a questionnaire.
- ③ Redesigning a survey (for factorization of final outcomes).
- ④ Interpreting the results.



9 Ideal Achievements

Q10. What are the impact evaluation goals?

1. To be accountable to the taxpayers.
2. To learn and implement cost-effective policies.

① Production of knowledge.

- Testing the new ideas/policies.
- Understanding each selection process.
- Estimating consistent treatment effects.
- Identifying the causal chains.
- Interpreting the results.

② Dissemination and use of knowledge.

- There is a small gap between the current and the ideal achievements.
- Are we close to attaining these impact evaluation goals?

Start



Goal

(we are here)

## 10 What to Do from Now?

Q11. How much and how should we, the donor and partner country community as a whole, produce the knowledge  $K$ ?

Principle: NPV (marginal benefits) = NPV (marginal costs).

① Nash equilibrium JBIC optimum: choose  $dK_{j,\tau}$  given  $K_{\tau-1}^{**}$ .

$$\sum_{\tau=t}^{\infty} \frac{V'_{j,K_j}(K_{j,\tau}^{**}, K_{-j,\tau}^{**})}{(1 + \delta_j)^{\tau-t}} = \sum_{\tau=t}^{\infty} \frac{C'_j(dK_{j,\tau}^{**})}{(1 + \delta_j)^{\tau-t}}.$$

JBIC is only considering itself in choosing the knowledge level  $K_{j,\tau}^{**}$ .

- ② Donor optimum: choose  $d\mathbf{K}_\tau$  given  $\tilde{\mathbf{K}}_{\tau-1}$ .

$$\sum_{\tau=t}^{\infty} \sum_{j=1}^I \sum_{i=1}^I \frac{V'_{j,K_i}(\tilde{K}_{j,\tau}, \tilde{\mathbf{K}}_{-j,\tau})}{(1 + \delta_j)^{\tau-t}} = \sum_{\tau=t}^{\infty} \sum_{j=1}^I \frac{C'_j(d\tilde{K}_{j,\tau})}{(1 + \delta_j)^{\tau-t}}.$$

Donors as a whole are only using their own resources.

- ③ World optimum: choose  $d\mathbf{K}_\tau$ , given  $\mathbf{K}_{\tau-1}^*$  and local context  $\mathbf{L}_\tau$ .

$$\sum_{\tau=t}^{\infty} \sum_{j=1}^I \sum_{i=1}^I \frac{V'_{j,K_i}(K_{j,\tau}^*, \mathbf{K}_{-j,\tau}^* | \mathbf{L}_\tau)}{(1 + \delta_j)^{\tau-t}} = \sum_{\tau=t}^{\infty} \sum_{j=1}^I \frac{C'_j(dK_{j,\tau}^* | \mathbf{L}_\tau)}{(1 + \delta_j)^{\tau-t}}.$$

Donors and partner governments are jointly using their resources.

- ▶ Optimal proportions  $C_j(dK_{j,\tau}^*)/(\text{total costs})$  can be different among the donors given the different cost structure,  $C_j \neq C_i$  for  $j \neq i$ . CGD.
- ▶ In principle, each donor should *follow its comparative advantage in impact evaluation* in a decentralized fashion, i.e., choose the impact evaluation particulars (type and location of projects) with large own marginal benefits  $V'_{j,K_j}$  with small marginal costs  $C'_j$ .
  - ▣▶ Imposing a fixed percentage rule may induce the donors to follow their comparative advantage, so it may not be so bad.

- 
- ▶ But the donors and the partner countries must *consult more closely* among themselves and with each other to exploit the positive externality in knowledge production  $V'_{j,K_i} > 0$ .
  
  - ▶ We should *use the country l's knowledge of local context*  $L_{l,\tau}$  more in impact evaluation, because:
    - It enhances the marginal benefits  $\frac{\partial V'_{j,K_i}}{\partial L_{l,\tau}} > 0$  by providing us with more accurate ways of measuring the impacts and more accurate interpretations of the causality.
  
    - It increases the marginal costs by less than the marginal benefit increase, if cost-effective local impact evaluators are employed.

- ▶ Another important increase in net marginal benefit can be obtained if *the donors work with the partner government in collecting the baseline* to come up with a candidate project list.

### REFERENCES

**Strauss, John, and Duncan Thomas (1998)**, “Health, Nutrition, and Economic Development,” *Journal of Economic Literature*, Vol. 39, Issue 2 (June), pp. 766-817.