

Development Research Group

Knowledge in Development Note 5

Assessing Results using Impact Evaluations

June 2007

Aid donors and governments want sound evidence on the impacts of their spending on poverty and human development. But because the design and implementation of sound evaluations of impact can be costly, they have been relatively rare in developing countries and, until recently, in World Bank programs. Instead, development policy has often been guided by methodologically flawed evaluations, anecdotal evidence, or development fashions.

Impact evaluations are essential to making aid more effective

Impact evaluations are about understanding what works in development, under what circumstances, and why. They are not about simply measuring inputs—such as how many school meals were delivered in a school feeding program. Instead, they are about assessing impacts on key development objectives, such as children’s nutritional status. By determining who gains from a program and by how much, impact evaluations are a key to making aid effective.

Impact evaluations contribute to a culture of accountability and results-based lending

Impact evaluations help foster a culture of accountability and results-based lending within the Bank and within client governments in developing countries. They cannot be used to evaluate the effects of every program or intervention—especially, where a policy is economywide. But credible impact evaluations should gradually become a central input to the design of effective policies.

Impact evaluations are essential for creating a “knowledge bank”

The number and quality of impact evaluations are gradually improving. More research is needed, but research also needs to move onto bigger questions. These include understanding why the same intervention can be successful in some circumstances and settings but not in others, and how to make policies more effective. Gaining this understanding will require making cost-effectiveness comparisons across policies. Are learning outcomes best improved through reducing poverty, building schools and other education-based policies, improving early childhood development through health interventions, or building rural roads, as advocated by infrastructure specialists?

Our understanding to date

Designing convincing impact evaluations involves constructing a credible comparison group

One reason rigorous evaluations have been rare is that measuring the impact of policies is difficult. Impact evaluation aims to compare outcomes for individuals who

participate in a program with the outcomes that would have obtained if they had not participated. Since they can never be observed, outcomes for participants in the absence of an intervention have to be proxied by those for a “comparison” group of similar individuals who did not receive the program. But any differences in characteristics between the two can result in misleading inferences about the program’s impacts. Three sets of techniques can be used to minimize such differences and construct ideal comparison groups for rigorous evaluations:

- *Randomization* involves selecting the beneficiaries of an intervention through a lottery. As in medical trials, potential beneficiaries are randomly divided into two groups—one receiving a program, another not. This is appealing because it effectively removes *all* differences between recipients and non-recipients, so that differences in outcomes between the two can be credibly interpreted as the impact of the program (rather than any confounding factors).
- Randomization is not always feasible, particularly when working with large government programs. And randomized evaluations can be deceptive about the gains from scaling up. With good data, it is possible to assess impacts without randomization. These *quasi-experimental methods* aim to remove all observable differences between beneficiaries and the comparison group of non-beneficiaries.
- If, as with randomization, planning for the evaluation begins early, it is possible to collect baseline and post-intervention data on participants and non-participants, and a number of *econometric methods* can be used to correct for differences between these two groups. For instance, participants can be carefully matched with non-participants on the basis of observable, pre-intervention characteristics. Alternatively, changes over time in outcomes for non-random samples of participants and non-participants can be compared. Often, the two techniques are combined.

Impact evaluations require planning and careful design at an early stage of a project, extensive coordination, and a substantial budget

To be rigorous and convincing, impact evaluations require careful design, generally before an intervention is put in place. They require coordination between researchers, others involved in project preparation, and government counterparts. They also require careful supervision to ensure that the evaluation design is respected during program implementation. And they may require sizeable budgets (for baseline and follow-up surveys). These are prerequisites for successful evaluations—even if they cannot assure that every evaluation will deliver the expected results.

Credible, rigorous impact evaluations have large returns because they advance our knowledge about the effects of policies, have lessons for the design of interventions, and help ensure the political sustainability of successful programs.

Credible evaluations highlight what interventions work, and why

In *Ecuador* and *Mexico*, cash transfers are made to poor households subject to a means-test. In both countries, random assignment among eligible communities and

families was used during the roll-out of the program, and this forms the basis of evaluations of the effect of these programs on the cognitive development of children of pre-school age. Both evaluations showed large, significant effects of cash transfers on cognitive development, in particular among the poorest households. Low levels of cognitive development in early childhood have very serious consequences for school performance and labor market outcomes, and these evaluations show that carefully targeted cash transfers are one of a menu of effective interventions.¹

An evaluation of the impacts on child health of piped water in *India* found lower prevalence and duration of diarrhea for children living in households with piped water as compared to a matched comparison group of households. Yet, it also found striking differences in the child health gains according to family income and adult female education. The health gains largely bypass children in poor families, particularly when the mother is poorly educated. These findings highlight the importance of combining public investments of this type of infrastructure with other interventions in education and income poverty reduction.²

In *Bangladesh* families who own more than one-half acre of land are ineligible for most group-based microcredit programs. This rule was used to compare the outcomes of families with “just under” and “just above” the maximum amount of land holdings. The evaluation showed large, positive effects of microcredit on household welfare. The results underscored the negative effects of credit constraints in developing countries and the potential of microcredit programs as a way of alleviating these constraints.³

In *Colombia* and *Cambodia*, demand-side incentives such as school vouchers and scholarships have been shown to have large, positive effects on enrollment in lower secondary school. In Colombia, children who received the voucher were 10 percentage points more likely to finish 8th grade, and were also more likely to perform well on tests of academic achievement. In Cambodia, poor girls who received scholarships were 30 percentage points more likely to attend lower secondary school than they would have been in the absence of the program.⁴

Chile Solidario, an innovative social protection system in *Chile*, targets households in extreme poverty and provides them with a two year period of psycho-social support through a local social worker. During this phase the households are ensured preferential access to a system of monetary transfers and a system of social programs that meet their needs in terms of human capital endowments, housing and income generation capacity. The evaluation exploits an exogenous geographic variation in the assignment of the program to find that the program improved education and health outcomes and increased take-up of cash assistance and social programs for housing and employment. The study provides suggestive evidence of the key role that close follow-up and psycho-social support has in enabling households to orient themselves towards the future and make the best use of public programs and services.⁵

In *China*, DECRG staff have been working with local counterparts to study the long-term impact of a major Bank lending operation that aimed to dramatically reduce absolute poverty in one of China’s poorest rural regions using multi-sectoral interventions (including farming, animal husbandry, infrastructure and social services).

Only modest gains to mean consumption emerged in the longer-term, but certain types of households gained more than others. The educated poor were under-covered by the community-based selection process — greatly reducing overall impact.⁶

Credible evaluations can have large effects on public policies

In *Mexico* PROGRESA, a conditional cash transfer program, made transfers to households if preschool children were taken for regular visits to health centers and school-aged children were enrolled in school. The program was phased in a gradual manner and, among eligible communities, a lottery was used to determine what areas would receive the program first. The randomized phase-in allowed researchers to convincingly show that PROGRESA improved child schooling and health. The PROGRESA evaluation resulted in the design of similar programs in almost every country in Latin America.⁷

In *China*, it is estimated that the poverty head count would be nearly 15 percent lower in the absence of out-of-pocket spending for health. However, an evaluation of an insurance scheme for catastrophic health shocks in urban areas showed no effect of the program on health spending. Part of the problem appeared to be that providers were paid on a fee-for-service basis, which encouraged them to shift insured patients from basic care to high-tech, high-margin services and drugs. To discourage providers from over-providing care, a World Bank-supported project in rural areas operated on both the demand-side by making the village-based rural health insurance system more attractive through subsidized premiums, and on the supply-side, by introducing treatment protocols, drug lists and training programs to reduce “demand inducement” by providers. An evaluation showed that the project significantly reduced catastrophic health spending.⁸

In *Argentina* matching techniques were used to estimate the impact of a program that provided support to families that had become unemployed during the severe economic crisis of 2002. The evaluation showed that the program reduced aggregate unemployment and extreme poverty. Together with earlier research on the impact of workfare programs, this evaluation has helped governments develop appropriate policies during macroeconomic shocks.⁹

Credible evaluations help ensure political sustainability

The rigorous nature of the PROGRESA evaluation was one of the reasons for the administration of President Fox to support a program inherited from the previous administration—an unusual outcome. The name of the program was changed (to *Oportunidades*), but the design was essentially left unchanged. The program has now continued with the administration of President Calderon. Much the same happened in Colombia, where a rigorously evaluated conditional cash transfer program survived an administration change—in part, because of the results of a credible impact evaluation.

Current and future research directions on impact evaluation

The development community has demanded better evaluation as an input to results-based aid. The World Bank—with its global scope, unparalleled access to policy-makers at the highest levels, and large volumes of lending for specific interventions—is

ideally placed to encourage learning about development from impact evaluations. Much can be learned from the accumulation of a large number of evaluations, on similar interventions, carried out in different settings. The results can then be aggregated into a review of what can be generalized about the circumstances under which policies work best, and the types of policies and design that best achieve a given gain for a given cost.

The Office of the Chief Economist and the Human Development Vice-Presidency has launched an important initiative to promote a large number of individual-level evaluations in various areas—including conditional cash transfers, provision of information to beneficiaries, school-based management, teacher contracting, early childhood development, and the provision of anti-retroviral drugs to AIDS patients.

Increasing the number and quality of evaluations of similar interventions in different circumstances will help make policies more effective

Building up the body of evaluations should also shed light about evaluation methods. Randomization may be the most convincing evaluation design, but it is not always possible or desirable. Understanding what other techniques work—under what circumstances, and what their biases are—is an important direction for future research.

Increasing the number and quality of evaluations will help our understanding of the techniques that are appropriate in different circumstances

Examples of ongoing research involving DECRG staff that aims to answer these questions include the following:

- *What policies work best to extend care and support to AIDS patients?* The HIV-AIDS epidemic is devastating the social fabric in many Sub-Saharan African countries. In addition to the human costs of the epidemic, the economic consequences are dire—families are condemned to poverty when breadwinners die, and communities struggle to cope with the burden of patients and orphaned children. A series of evaluations of the impact of provision of antiretroviral drugs are being conducted in *Burkina Faso, Ghana, Mozambique, Kenya, Rwanda and South Africa*.
- *Can increased school-based management improve learning outcomes?* It is often argued that decentralizing decision making in education from central ministries to school-based local management committees of parents and school administrators can result in improvements in learning outcomes. An ongoing randomized evaluation of school-based management in *Nepal* directly tests the impact of policies that transfer responsibilities to local management committees.
- *Do new roads linking poor isolated rural areas to the outside world promote local development and higher living standards?* Despite a general consensus on the importance of rural roads, there is surprisingly little hard evidence on the size and nature of these benefits and on the contextual factors that influence outcomes. Ongoing evaluations in both *Vietnam* and *Bangladesh* examine these issues, focusing on the heterogeneity in impacts, and the nature of that heterogeneity,

including the interactions with geographical, community and household characteristics.

- *To what extent can school fee reductions improve learning outcomes?* The extent to which fees paid by households are a significant deterrent to learning outcomes in developing countries is a question of great policy importance. An ongoing quasi-experimental evaluation in *Lesotho* studies the impact of a nationwide policy that eliminated most fees at the primary school level.
- *What are the gains from iron deficiency treatments on health and productivity outcomes?* Micronutrient deficiencies, widespread in many developing countries, can have very large costs in terms of health, schooling, and wage-earning potential. An ongoing randomized evaluation in *Indonesia* tests the effect of iron supplementation on health, productivity, and wages.

[Contact: Dominique Van De Walle, Dvandewalle@worldbank.org, 202-473-7935]

Endnotes

Most Bank documents cited in this summary are available through the documents and reports portal of the World Bank <http://www-wds.worldbank.org/>. The word “processed” describes informally reproduced works that may not be commonly available through library systems.

¹ On Ecuador, C. Paxson and N. Schady. 2007. “Cognitive Development among Young Children in Ecuador: The Roles of Wealth, Health, and Parenting.” *Journal of Human Resources* 42(1): 49-84; and C. Paxson and N. Schady. 2007. “Does Money Matter? The Effect of Cash Transfers on Child Health and Development in Rural Ecuador.” World Bank Policy Research Working Paper 4226.

On Mexico, see L. Fernald, P. Gertler, and L. Neufeld. 2006. “How Important is the Amount of Cash in Conditional Cash Transfer Programs for Child Development?” University of California at Berkeley. Processed.

² J. Jalan and M. Ravallion. 2003. “Does Piped Water Reduce Diarrhea for Children in Rural India?” *Journal of Econometrics* 112: 153-173.

³ M. M. Pitt and S. Khandker. 1998. “The Impact of Group-Based Credit Programs on Poor Households in Bangladesh: Does the Gender of Participants Matter?” *Journal of Political Economy* 106(5): 958–96.

For a different perspective see J. Morduch. 1999. “The Microfinance Promise.” *Journal of Economic Literature* 37(4): 1569–1614.

⁴ On Colombia, see J. Angrist, E. Bettinger, E. Bloom, E. King, and M. Kremer. 2002. “Vouchers for Private Schooling in Colombia: Evidence from a Natural Randomized Experiment.” *American Economic Review* 92(5): 1535-58.

On Cambodia, see D. Filmer and N. Schady. 2007. “Getting Girls Into School: Evidence from a Scholarship Program in Cambodia.” *Economic Development and Cultural Change* (forthcoming)

⁵ E. Galasso. 2006 “With their effort and one opportunity: Alleviating extreme poverty in Chile,” Development Research Group, World Bank, Washington, DC, processed.

⁶ S. Chen, R. Mu and M. Ravallion, 2006 “[Are there lasting impacts of aid to poor areas? Evidence from Rural China](#),” Policy Research Working Paper 4084.

⁷ P. Gertler. 2004. “Do Conditional Cash Transfers Improve Child Health? Evidence from PROGRESA’s Control Randomized Experiment.” *American Economic Review* 94(2, Papers and Proceedings): 336–41; T. P. Schultz. 2004. “School Subsidies for the Poor: Evaluating the Mexican PROGRESA Poverty Program.” *Journal of Development Economics* 74 (1): 199–250.

⁸ A. Wagstaff, and S. Yu, “Do health sector reforms have their intended impacts? The World Bank’s Health VIII Project in Gansu Province, China,” *Journal of Health Economics* 26(3): 505–35

⁹ E. Galasso and M. Ravallion. 2004. “Social Protection in a Crisis: Argentina’s Plan Jefes y Jefas.” *World Bank Economic Review* 18(3): 367–99.