

TOWARD CLOSING THE EVALUATION GAP: LESSONS FROM THREE RECENT IMPACT EVALUATIONS OF SOCIAL PROGRAMS IN LATIN AMERICA AND THE CARIBBEAN*

Lorenzo Moreno

Mathematica Policy Research, Princeton, N.J.
lmoreno@mathematica-mpr.com

Larissa Campuzano

Mathematica Policy Research, Princeton, N.J.
lcampuzano@mathematica-mpr.com

Dan Levy

John F. Kennedy School of Government, Harvard University, Cambridge, MA
dan_levy@harvard.edu

Randall Blair

Mathematica Policy Research, Princeton, N.J.
rblair@mathematica-mpr.com

Abstract

Despite recent growing demand from funders and governments, rigorous impact evaluations in Latin America and the Caribbean remain the exception rather than the rule. Many commissioned impact evaluations are methodologically weak, and thus only marginally useful

*The views expressed here are those of the authors and should not be attributed to Mathematica Policy Research, the Millennium Challenge Corporation, the Inter-American Development Bank, the World Bank, or any of these agencies' member countries.

Acknowledgments: Several people provided valuable suggestions for making this paper more useful. Rebecca H. Tunstall of the Millennium Challenge Corporation (MCC) encouraged us to write this document and made insightful comments about MCC's policies and El Salvador Compact. Vince Ruddy, MCC regional director in El Salvador, and Franck Wiebe, MCC chief economist, also provided valuable comments and clarifications that improved this paper. At Mathematica Policy Research, our colleague Kenneth Fortson reviewed this paper and provided useful suggestions for improving it. In addition, we received useful observations from Mathematica colleagues at an internal seminar. We would also like to thank Iliana Yashine at *El Colegio de México* in Mexico, for her detailed review of our paper and her insightful observations. Also, we are indebted to the participants of the conference "*Haciendo Operativas las Evaluaciones de Impacto de Programas Sociales*," held in Mexico City on February 13, 2009. Finally, we are grateful to Ernesto Treviño and María José Ramírez, at *Universidad Diego Portales* in Chile, for their useful comments on this paper.

in assessing the impact of social interventions. Other impact evaluations feature strong research methodologies at their conception, but face considerable institutional challenges during key points in the design and implementation phases. This paper identifies some of the barriers that limit the design and implementation of rigorous impact evaluations in this region, as well as several enablers to the successful design and implementation of such evaluations. The paper also outlines some key practices for designing and implementing high-quality impact evaluations in Latin America and the Caribbean. We use a case study methodology that combines our experience designing and implementing impact evaluations in three ongoing or recent social programs in El Salvador, Jamaica, and Mexico.

— Key words: Impact evaluation, evaluation gap, best practices, human development programs, productive development programs, case study, institutional factors, Latin America and the Caribbean.
JEL classification: C93, H43, O19.

Introduction

After five decades of work and billions of dollars invested by governments of developing countries and multilateral institutions, we have a limited understanding of what types of interventions are successful in alleviating poverty and improving education, employment, health, and nutrition outcomes in developing countries. Impact evaluations—studies that seek to assess whether changes in social and economic outcomes can be attributed to specific programs—have the potential to substantially enhance such understanding. By measuring the effect of social programs with a high degree of accuracy, reliability, and validity, impact evaluations can increase development assistance organizations’ knowledge base and provide them with a road map for future interventions (Savedoff and Levine 2006).

The most rigorous approach to impact evaluation is random assignment, by which some units (such as individuals, groups, or geographic areas) are randomly assigned to an intervention and other units are randomly assigned to the control group (that is, the *statu quo*). The strength of random assignment is that it relies on only a few assumptions to assert that an intervention caused the measured results. For this reason, random assignment is widely accepted among academics, government agencies, and multilateral funding institutions as the “gold standard” of impact evaluation.¹ However, random assignment is not appropriate for every impact evaluation—particularly in cases of limited generalizability of findings to other populations or when serious ethical considerations or logistical barriers prohibit limiting potential beneficiaries’ access to services (Ravallion 2009). In these instances, evaluators can employ quasi-experimental designs and a variety of statistical techniques to substitute for random assignment.

¹ For example, the U.S. Department of Education’s What Works Clearinghouse, which reviews and assesses numerous education studies at a national level, considers random assignment the optimal evaluation design. The highest rating a well-implemented quasi-experimental design can receive from the Clearinghouse is “*Meets Evidence Standards with Reservations*”.

Although development assistance organizations recognize the merits of rigorous impact evaluation, they often face incentives that discourage the widespread use of these evaluations. First, an impact evaluation is a long-term investment in the context of short-term project cycles. The benefits of an impact evaluation can take years to materialize, while its costs are immediate and often substantial. Second, rigorous impact evaluations require coordination and sometimes modifications of the implementation plans. Third, impact evaluations introduce a new degree of accountability to the development assistance process for both funders and implementers. As such, program staff and managers often see impact evaluation as a threat—especially if results are not positive (Pritchett 2002). Consequently, donors and national governments traditionally do not demand or produce enough rigorous impact evaluations, thus creating an evaluation gap (Levine 2006).

Several international nongovernmental organizations, such as the Center for Global Development, have actively advocated for closing the evaluation gap (Savedoff, Levine, and Birdsall 2005). Through their public testimony, studies, and reports, these proponents of impact evaluation have gained many allies in the United States Congress and major development institutions, such as the World Bank and the United Nations. Their cause has been further strengthened by recent advances in economic and statistical methods for impact evaluation, which have catalyzed acceptance of evaluation as a tool for effective policymaking and program management (Duflo, et al. 2006). Furthermore, this cause has been supported by rigorous evaluations of poverty-reduction programs in Latin America and the Caribbean, such as *PROGRESA-OPORTUNIDADES* in Mexico and *Bolsa Familia* in Brazil, which are winning recognition worldwide (Treviño 2008).

As a result of these advocacy efforts and growing acceptance of rigorous social research, demand for impact evaluation has incrementally grown in recent years—particularly in Latin America and the Caribbean. Funders, primarily international and regional development banks and foundations, have begun to require impact evaluations as part of their loan or grant requirements. In addition, several national and regional governments in this region now require impact evaluations of social programs to enhance accountability and transparency of their policies, make decisions about public spending allocation, and assess program scalability and sustainability. For example, Mexico passed public administration reform legislation in 2007 regulating the evaluations of federally funded social programs (Estados Unidos Mexicanos 2007). This legislation culminated a process started at the beginning of the current decade that required impact evaluations of these programs, among other types of evaluations (CONEVAL 2009).

Despite this recent growing demand on the part of funders and governments, high-quality and rigorous impact evaluations of government programs in Latin America and the Caribbean remain the exception rather than the rule.² Many impact evaluations that are commissioned are

² For instance, among informal skills development programs, only two evaluations have used experimental methods to assess the impacts of subsidized, basic skills and technical/vocational training programs serving low income youths in Colombia and the Dominican Republic (Attanasio et al. 2008; Card et al. 2007). Likewise, researchers in Chile only recently began considering a rigorous evaluation of an intervention to enhance teacher quality for preschool children (Un Buen Comienzo 2008), despite broad interest in evaluating early childhood interventions in that country. Examples of high-quality impact evaluations of other social interventions in the region are becoming more common but are still rare.

methodologically weak, and thus only marginally useful in assessing the impact of social interventions. This is mostly the result of misconceptions among both the funders and the program managers about the potential consequences of using rigorous designs for impact evaluation, as well as a lack of an evaluation culture that appreciates the advantages of rigorous impact evaluations or, in the extreme, of evaluation at all. Other impact evaluations feature strong research methodologies at their conception, but face considerable institutional challenges and barriers during key points in the design and implementation phase, such as a lack of support from program implementers, hasty implementation schedules, or scarce financial and human resources to sustain the impetus of the long-term operations involved in conducting a rigorous impact evaluation.

This paper identifies barriers to, as well as enablers of, successful implementation of high-quality and rigorous impact evaluations in Latin America and the Caribbean. For this purpose, we use a case study methodology that combines our experience designing and implementing impact evaluations in three ongoing social programs in El Salvador, Jamaica, and Mexico. Although these barriers and enablers are particularly salient in Latin America and the Caribbean, we believe they apply in varying degrees to low- and high-income countries alike.

The remainder of this paper is organized as follows: First, we describe the context of the evaluations. Second, we identify and discuss the barriers to effective impact evaluation. Third, we identify and discuss enablers of effective impact evaluation. Fourth, we outline some key practices based on our experience in the expectation that they may assist other evaluators to more effectively design and implement rigorous impact evaluations. Finally, we discuss the evaluators' desideratum in the context of implementing high-quality impact evaluations in Latin America and the Caribbean.

1. Context of the Three Evaluations

We conducted or are currently in the process of conducting impact evaluations of the following three programs: (1) two major components of the Millennium Challenge Corporation (MCC) "Compact" for El Salvador; (2) the "Program for the Advancement Through Health and Education" (PATH), implemented in Jamaica; and (3) the infrastructure component of *Hábitat*, implemented in Mexico.

The objective of all three programs is to reduce poverty in targeted regions of the respective countries (Table 1 of the Annex). However, the programs are fundamentally different in their scope and primary activities. In El Salvador, contractors will provide technical and material assistance in the areas of education and productive development in an underdeveloped region (Northern Zone), which was the locus of the country's civil war in the 1980s (Millennium Challenge Corporation 2008). Under Jamaica's PATH program, the government provides conditional cash transfers to participant households to promote investment in health and education of their children and high-risk adults (Levy and Ohls 2007). In Mexico's *Hábitat* program, local governments receive funds to support a variety of social projects (Campuzano and Levy 2007). Specifically, the programs target marginalized or poor individuals or producers in both urban and rural areas (El Salvador); poor children, pregnant women, lactating mothers, the disabled, and the elderly (PATH); and marginalized

families in large metropolitan areas (*Hábitat*). Regarding their duration, the *Hábitat* and PATH interventions are open-ended; El Salvador's MCC Compact is scheduled to run for five years.

In the El Salvador program, the executive branch, with support from the legislature, established a public agency, the *Fondo del Milenio (FOMILENIO)*, to hire implementing organizations and manage the program's various components. In the PATH and *Hábitat* programs, existing government ministries handle program implementation. All three programs represent a substantial portion of their country's social investments. At \$461 million U.S. dollars, the El Salvador program constitutes an especially large monetary investment, equivalent to 1.4 percent of El Salvador's gross domestic product (GDP) in 2006 (Table 1).

Funding for the *Hábitat* and PATH programs was provided by a regional development bank (Inter-American Development Bank) and an international development bank (the World Bank), respectively (Table 1). Funding for the El Salvador Compact was provided by a U.S. development agency (MCC) through an innovative aid program that encourages country ownership (Phillips-Mandaville 2008). In all three programs, the funding institution mandated an impact evaluation, set aside funds for it, and imposed legal requirements for the evaluation. Noteworthy is MCC's policy of aid effectiveness, which looks for opportunities to use randomized designs or the next most rigorous alternative as part of its monitoring and evaluation oversight of its compacts and other programs (Wiebe 2008).

We completed impact evaluations for the PATH and *Hábitat* programs in 2007, and will complete an impact evaluation for El Salvador's Compact in 2012. The entities contracting these evaluations were MCC for the El Salvador Compact evaluation; the government of Jamaica for the PATH evaluation; and the government of Mexico for the *Hábitat* evaluation. There was a formal procurement process for all three contracts. We were accountable to the Ministry of Labour and Social Security (MLSS) and *Secretaría de Desarrollo Social (SEDESOL)* for the Jamaica and Mexico evaluations, respectively. We are accountable to MCC for the El Salvador evaluation (Table 1).

Mexico has some experience with impact evaluations of social programs, stemming from the internationally recognized impact evaluation of its flagship social program, *PROGRESA-OPORTUNIDADES* (Levy 2006), and the subsequent legislation requiring *SEDESOL* and other government agencies to evaluate all programs. Most recently, Mexico created a government entity that is responsible for the evaluation of its social programs (*Consejo Nacional de Evaluación de la Política de Desarrollo Social*, or *CONEVAL*), so its experience with impact evaluation is growing rapidly. In contrast, El Salvador and Jamaica have very limited experience with impact evaluation. Furthermore, the implementing organizations in El Salvador and Jamaica had only narrow experience with impact evaluation.

Overall, the El Salvador Compact, PATH, and *Hábitat* programs rely on innovative and ambitious social interventions to reduce poverty in marginalized zones. The funding agencies mandated independent impact evaluations of their sponsored programs. Given the relatively limited experience with impact evaluation in each of the three countries, it is not surprising that there were substantial barriers to their implementation. We discuss these barriers in detail in the following section.

2. Barriers to Implementing a Rigorous Evaluation

Despite the initial enthusiasm and excitement of funders, program managers, and the independent evaluator regarding the design and implementation of rigorous impact evaluations of the El Salvador Compact, PATH, and *Hábitat* programs, the following four institutional barriers complicated these tasks:

2.1 Lack of support for rigorous evaluation designs

In El Salvador and Jamaica, senior political leaders and program managers had important concerns about the equity and fairness of allocating beneficiaries to the interventions using random assignment. For the El Salvador evaluation, MCC had initially encouraged the use of random assignment where feasible and we had proposed it for several Compact interventions. Nevertheless, due to the skepticism of *FOMILENIO* and the local implementer about the practicality of randomized designs, this type of design was eventually rejected (except for the scholarship component, in which *FOMILENIO* agreed to introduce randomization if demand for scholarships by eligible candidates exceeded supply).³ The arguments of *FOMILENIO* and the local implementers in favor of discarding experimental designs included the belief that random assignment of individuals was unethical and largely infeasible, and that this methodology could delay the implementation of program services, particularly in the productive development program. *FOMILENIO* also argued against randomized designs because its intent was to select “winners” (that is, beneficiaries that would truly benefit from the intervention, resulting in large program impacts), but it was unsure that there would be enough of them at the beginning of the compact to warrant their random assignment into intervention and control groups.⁴ In Jamaica, random assignment also was judged to be operationally infeasible and costly, so support for an experimental design quickly waned. In Mexico, random assignment designs were not feasible given the stage of implementation of *Hábitat*.

³ However, recently, an agreement among *FOMILENIO*, Mathematica, and the local implementor was reached to use random assignment for the evaluation of the Productive Development Project.

⁴ This was a valid concern, as *FOMILENIO* was unable to reach compact-mandated targets for program enrollees in other components of the El Salvador Compact intervention. MCC noted that, when using the cost-benefit ratio as its primary selection criteria for potential beneficiaries of the water and sanitation components, which favored proposals with the greatest beneficiary numbers per dollar spent, the resulting portfolio of potential beneficiaries narrowly missed the full number of beneficiaries contemplated in the Compact. However, had randomized selection been used, the average cost-benefit ratio would have been significantly lower, meaning an even more significant breach of the goals defined and agreed between *FOMILENIO* and MCC in the Compact. The Compact’s status as an international treaty raised the importance of this drawback of using experimental designs with the selection process.

Despite our intense and sustained efforts to allay the concerns of both El Salvador Compact and PATH leadership regarding experimental designs, these designs were not adopted because of issues related to their political, logistical, and financial feasibility. Most of these issues could have been addressed if the leadership had been ready to address the restrictions, some of them minor, that experimental designs involve. Thus, although a core group of program managers in Jamaica and El Salvador were committed to rigorous impact evaluation, support from senior management—including senior government officials—was not clearly available. This limited support constrained our ability to implement experimental designs for these two impact evaluations.

2.2 Ambitious and unrealistic schedules for designing and implementing programs

Another factor that complicated the use of rigorous evaluation designs was the timing of the evaluation relative to the start of program operations. For instance, in Mexico, the *Hábitat* program had already begun when we started designing the impact evaluation. Because randomization of eligible areas was no longer feasible at that time, our only option for the evaluation design was to propose a quasi-experimental design (Table 2 of the Annex). Furthermore, we were not able to collect baseline data and had to rely on secondary (census) data to measure outcomes on intervention and comparison areas before and after the start of the demonstration program. In Jamaica, the evaluator was selected at nearly the same time the program was launched, but it took the government many months to issue an evaluation contract and begin working with the evaluator. The proposed randomized design relied on providing program services to households that were nearly (but not quite) eligible for the program. The El Salvador evaluation started before the programs were implemented. In theory, this would have made it possible to conduct key components of a high-quality evaluation, such as collecting baseline data before the start of the programs. However, *FOMILENIO* selected the implementing contractors after the evaluator had been selected, and this made coordination among all entities regarding the evaluation quite challenging. This decision also complicated the process of selecting the best possible design because the implementers' compensation was tied to their performance implementing the program on schedule. Therefore, the urgency to get program operations going as soon as possible did not mesh well with the schedule required to implement certain designs, such as random assignment.

2.3 Missed opportunities to build on lessons learned

Although the El Salvador Compact and PATH programs had a pilot program, the opportunity to build on these pilots to develop an evaluation design jointly with the development of the program did not fully materialize (Table 2). Furthermore, none of the early pilots were built into the program from the beginning or were slated for formative (not impact) evaluation. For instance, in El Salvador, *FOMILENIO* was urged by MCC to expedite implementation of pilot activities based on existing field-based organizations to generate early results and associated lessons learned. This coincided with pressure felt by both *FOMILENIO* and MCC to more quickly deliver assistance and disburse

funds to beneficiaries. As a result, FOMILENIO's pilot test of the productive development intervention did not generate a full understanding of the target population, the key research questions, nor the process for implementing the interventions—including the recruitment of potential participants. This delayed the development of an optimal impact evaluation design and exerted considerable pressure on the design and collection of baseline data. In Jamaica, the pilot program was studied and, although some useful lessons were drawn for the purposes of full-scale program implementation, the study was not as useful for the purposes of impact evaluation design.

2.4 Limited feasibility of using existing secondary data

Another area of contention was the quality and timeliness of secondary data. In Mexico, as noted, only secondary data were available for the evaluation. Although these data were sufficient in describing communities' access to drinking water, sewage, drainage, and electricity, they limited the scope of the evaluation to service access. The use of census data also limited the evaluation time frame to two points in time. Furthermore, in all three countries, secondary data were typically limited to census data (population and/or economic) and one or two national household surveys (Table 2). Most of these data were of reasonable quality, although they were not representative of the regions where the programs were implemented and they were unavailable for the time periods required for measuring baseline or follow-up outcomes.

3. Enablers to Implementing a Rigorous Evaluation

Although barriers to rigorous impact evaluation in the region were formidable for these three programs, several factors contributed to the accuracy, reliability, and validity of these impact evaluations. The following factors served as enablers of rigorous impact evaluation of the El Salvador Compact, PATH, and *Hábitat* programs.

3.1 Support from program officers

All three evaluations benefited from capable and conscientious program officers within implementing or funding agencies who were dedicated primarily to monitoring and evaluation tasks. Particularly notable is the MCC program officer in charge of monitoring and evaluation of the El Salvador Compact, who has secured data, advocated for random assignment designs, and mediated among MCC, implementers, and evaluators throughout the course of the evaluation. In addition, *SEDESOL* and *MLSS* program staff helped refine study designs and obtain data for the *Hábitat* and *PATH* evaluations, respectively. In all three evaluations—and in the El Salvador

Compact evaluation in particular—program officers’ contributions were pivotal in building support for strong evaluation designs, securing these designs over non-rigorous alternatives, and providing the data necessary to conduct a rigorous impact evaluation.

3.2 Availability of funding for independent, high-quality primary data collection

Fortunately, in both Jamaica and El Salvador funding was made available for primary data collection. In the case of Jamaica, three independent household surveys were conducted for the purpose of the evaluation. In El Salvador, MCC funded a redesign of one of the annual national household surveys so that estimates from the expanded survey would be representative of the Northern Zone and a few departments within that region. We plan to use these new data in our evaluation of the productive development program to the extent possible. Furthermore, MCC is funding several rounds of independent surveys for evaluating the productive development, water and sanitation, and infrastructure components. These independent, high-quality primary data will yield valid, reliable, and accurate estimates of the impact of the intervention.

3.3 Existence of a policy and legal framework with political support

The impact evaluation of the El Salvador Compact was unique in that it featured a well-defined policy and legal framework governing all implementation and evaluation activities. The MCC Compact between the U.S. and Salvadoran governments defined the objectives of program and evaluation activities; designated responsibilities among all key actors; and mandated that the most rigorous impact evaluation designs be implemented, if feasible. As a result, nearly all stakeholders in the evaluation had a basic understanding of their roles, as well as the range of acceptable evaluation design options from the outset. However, during the design stage of the evaluation some *FOMILENIO* program managers expressed their preference for non-rigorous designs. Citing the Compact’s mandate for rigorous evaluation, the MCC program officer was able to influence these managers to incorporate stronger evaluation designs into their program operations. If this legal framework had not been in place prior to program implementation, there is no guarantee that negotiations with program managers would have resulted in rigorous evaluation designs.

4. Key Practices

In this section, we present five key practices for designing and implementing high-quality impact evaluations in Latin America and the Caribbean. These practices can be employed both to overcome many of the institutional barriers discussed in Section 2, as well as to harness the potential benefits of enablers discussed in Section 3. The key practices include:

4.1 Maintain early communication with key decision makers to identify and mitigate potential barriers to implementing a high-quality impact evaluation

Ideally, several months before program implementation is planned to begin, evaluators should meet with the funder's project officer, government officials of the implementing country, and program managers involved in the evaluation. In these preliminary meetings, evaluators can gauge stakeholders' knowledge and attitudes concerning evaluation, discuss potential designs, and identify key institutional barriers to rigorous evaluation—particularly issues of programmatic support and timing of the intervention and/or evaluation. When barriers are identified, evaluators can discuss with stakeholders the concerns that underlie these barriers, and take appropriate steps to mitigate potential threats to a rigorous study design. This practice is particularly salient if a pilot phase is being considered and the additional time required to introduce new concepts is available. Although conducting a full impact evaluation of the pilot phase is not realistic, typically due to limited sample size and time limitations on programs with relatively short periods of implementation, the (formative) evaluation of such a pilot would offer key stakeholders the opportunity to better understand the main components of an evaluation design, which are: (1) target population, (2) selection process of potential participants, (3) key research questions, (4) nature of the intervention, (5) key outcomes, and (6) feasibility of collecting baseline and follow-up data. To the extent possible, this evaluation of the pilot should be included in the overall evaluation budget. Thus, the evaluator will be in a better position to design an impact evaluation that is adequately customized to the needs of the funder, the program, and the implementing country.

4.2 Communicate to program managers that the evaluation will provide answers to their questions

Much of the lack of support for rigorous evaluation designs is based on a misunderstanding of the primary objectives and uses of impact evaluation. If evaluators can successfully communicate to program managers that the evaluation can address some of their fundamental questions about the intervention—as well as explain exactly how and when the evaluation will answer these questions—they can build much-needed support for rigorous designs. Evaluators should communicate the pros and cons of each potential evaluation design in a clear and concise manner, as well as spend

time with program managers and other stakeholders discussing how the impact evaluation will address the country's policymaking process. This would include discussing technical capacity building and opportunities to enhance or expand the data collected for evaluating other social programs. Communicating these ideas is another reason that it is critical for evaluators to meet with program managers well in advance of program implementation.

4.3 Develop evaluation designs that will not disrupt program implementation

A great way to generate support among program managers for rigorous, high-quality impact evaluations is to develop a design that will not complicate program implementation or burden program staff with additional responsibilities. This is a well-founded concern that program managers typically raise, because any delays or disruptions to their programs can have substantial economic and political consequences (and might even put their jobs at risk). In addition to working out with program managers the schedule for randomization of individuals, groups, geographic areas, or other intervention units as part of potential beneficiary selection, evaluators might want to build on the evaluation of the pilot phase to fine-tune their design. Although this exercise may add a few weeks to the evaluation design process and, as noted, require additional resources, it might be invaluable for demonstrating to program managers that the evaluation will not delay or derail the full-scale implementation of the program.

4.4 Limit the discussion of the technical aspects of the designs to a small group of interested stakeholders

Involving a wide array of stakeholders in discussions and negotiations related to study designs only complicates an already intricate process. Misinformation raised in these discussions can also erode support for rigorous designs. A good strategy is to involve only key stakeholders—funders, management, and principal consultants—in discussions of the technical aspects of study designs. This streamlines the study design process, increases the likelihood that stakeholders will settle on a rigorous design, and facilitates buy-in among key parties involved in the evaluation. Subsequently, the funders, program managers, or evaluators can brief senior government officials so that they can approve or address broad issues that require their attention without having to spend time and effort understanding the technical aspects of an evaluation design. Achieving a balance between transparency and stakeholder involvement, which are core values of funders such as MCC, and widespread discussion of technical issues poses a challenge to evaluators. To address this challenge, they should tactfully work with the project officer to determine whom to involve in key design and technical discussions and when to do so.

4.5 Ensure that all parties understand and comply with the legal framework for funding and implementing the evaluation

If there is considerable resistance from key stakeholders regarding the use of rigorous designs or ensuring the conditions for conducting a high-quality evaluation, evaluators should remind all parties of the mandates and contract stipulations governing the impact evaluation. To do so, evaluators need to fully understand the funder's mandate for impact evaluation, communicate regularly with the funder's project officer, and request his or her intervention when there is a high likelihood that a suboptimal impact evaluation design might be chosen.

Factors that could assist and enhance these practices include the evaluator's fluency in the main languages spoken in the country under consideration; full understanding of the culture, history, and politics of the country under consideration; the ability to adapt to rapid and frequent changes in circumstances; and tactful and effective communications with senior officials and technical and management staff.

5. The Evaluators' Desideratum

The key goal of any impact evaluation should be to provide stakeholders with credible estimates of the impacts of a program. On the one hand, the role of evaluators is to propose the most rigorous evaluation design to maximize the credibility of the estimates. On the other hand, the ultimate choice of evaluation design frequently arises from a negotiation process that includes logistical, political, and financial considerations. The evaluator's role in this process is to present the tradeoffs among competing evaluation designs clearly and to try to persuade the key decision makers of the value of rigorous designs.

In all three cases presented here, the most rigorous design was not selected because it was deemed infeasible from a logistical, political, or financial perspective. In all three programs, we and the contracting entities chose quasi-experimental designs or case study methodologies following a series of negotiations among stakeholders. However, these negotiations were far from straightforward and facile. For instance, in El Salvador, MCC had to exert considerable pressure on FOMILENIO's leadership to ensure that baseline data collection for several components was not delayed after the start of program operations or scrapped entirely. FOMILENIO's leadership at the time understood the importance of starting data collection early, so that appreciation of full—rather than partial—program impacts could be generated. However, the relatively weak capacity of the organization during its first year of establishment and operations, combined with competing priorities, led to delays of several months. However, this has not compromised our ability to undertake an impact evaluation of mixed rigor. In Mexico, we decided to withdraw from evaluating an extension of the *Hábitat* program when program administrators at *SEDESOL* moved to impose their own methodology for measuring key outcomes and appointed a non-independent contractor to conduct data collection. In Jamaica, we presented design options to the key stakeholders and

ultimately settled on a design that was methodologically strong but not as advantageous as the experimental design initially considered.

Thus, as evaluators, we walk the fine line between accepting less rigorous designs than desired or feasible and the mandate to produce credible and solid evidence of the impacts of the programs. Despite these challenges and the institutional barriers discussed above, in our experience, the enthusiasm and support of the project officers of the funding agencies and a core group of technical staff in each of the three countries facilitated the negotiation of compromises that, although not ideal, still enabled us to use fairly rigorous designs in evaluating the programs in these countries.

In sum, whether combined or individually, the key practices discussed in the preceding section can facilitate the process of designing and implementing high-quality impact evaluations in Latin America and the Caribbean. At a time in which reliable answers to the effectiveness of social programs are critical for governments and funders to move forward with and sustain social policy reform, every effort to reduce barriers to rigorous and high-quality impact evaluations by relying on the best evaluation practices is a much-needed step in the right direction.

Annex
Table 1
Context for the Impact Evaluations of Three Social Programs in El Salvador, Jamaica, and Mexico

Features	Program		
	MCC Compact (El Salvador)	PATH (Jamaica)	Habitat (Mexico)
Goals	Reduce poverty in an underdeveloped region (Northern Zone) by investing in human capital formation and supporting productive development activities, among other activities	Reduce current poverty by providing monetary transfers; reduce future poverty by encouraging poor households to invest in the health and education of their children	Reduce urban poverty by improving infrastructure, community development, and social services
Type of intervention ^{1/}	Technical and material assistance in the areas of education, training, and productive development	Conditional and unconditional cash transfers to participant households	Funds allocated to state and city governments to support urban development and risk prevention, urban planning, and social and community development
Number of intervention components ^{1/}	Multiple: Two large-scale projects, “Human Development” and “Productive Development”, each with several components	Two: Conditional cash transfers for children and unconditional cash transfers for the elderly	Multiple potential components, but participating communities choose an intervention according to their needs
Target population ^{1/}	Productive Development Project: poor individuals and organizations that benefit poor people in El Salvador’s Northern Zone Human Development Project, technical high school component: poor students in the most needy technical high schools in the Northern Zone Human Development Project, scholarship component: deserving, poverty-stricken youths in the Northern Zone	Child assistance grants: health grants to poor children up to age 6, and education grants to poor children ages 6 to 17 Social assistance grants to adults: poor, pregnant women or poor, lactating mothers; elderly poor (over age 65); and poor, disabled, and destitute adults under age 65	Marginalized families in cities or metropolitan areas with more than 15,000 inhabitants

Table 1 (continued)

Features	Program		
	MCC Compact (El Salvador)	PATH (Jamaica)	Hábitat (Mexico)
Duration	Five years: 2007-2012	Open ended	Open ended
Timing of intervention	All components concurrently implemented	All components concurrently implemented	Most components concurrently implemented
Institution responsible for implementing the program	Semiautonomous public agency	Government ministry	Government ministry
How the program fits in the country's social policy	Total funding for all MCC projects constituted 1.4 percent of El Salvador's GDP in 2006	PATH replaced three previous social programs to become the country's major income support program; PATH expenditures represented less than 0.2 percent of Jamaica's GDP in 2008	One of 12 social programs currently implemented by SEDESOL, Hábitat expenditures represented less than 5 percent of SEDESOL's budget in 2008, and less than 0.1 percent of Mexico's GDP in 2008
Funding level	\$478 million U.S. dollars from 2007-2012 for all MCC activities in El Salvador ^{2/} Productive Development Project: \$59 million for production and business services and \$22 million for investment support ^{3/} Human Development Project: \$21 million for formal technical education (including scholarships) and \$3 million for technical assistance ^{4/}	\$86.5 million U.S. dollars from 2001-2005; a \$44.6 million loan from the World Bank; and \$41.8 million from the Government of Jamaica, according to initial agreement signed with the World Bank ^{5/}	\$1.3 billion U.S. dollars from 2003-2006 ^{6/}
Type of funding	Grant	Loan	Loan
Funding source	U.S. government agency (Millennium Challenge Corporation)	World Bank	Inter-American Development Bank

TOWARD CLOSING THE EVALUATION GAP:
LESSONS FROM THREE RECENT IMPACT EVALUATIONS OF SOCIAL PROGRAMS IN
LATIN AMERICA AND THE CARIBBEAN

Table 1 (continued)

Features	Program			Habitat (Mexico)
	MCC Compact (El Salvador)	PATH (Jamaica)		
Impact evaluation required by funding agency	Yes	Yes	Yes	Yes
Set-aside funds for impact evaluation	Yes	Yes		Yes
Legal requirements for the evaluation	Yes	Yes, strongly supported by the government		Yes
Organization responsible for the evaluation	Independent, selected by funder	Independent, selected by local government under World Bank procurement rules		Independent, selected by funder
Hiring organization	U.S. government agency: MCC	Government of Jamaica: Ministry of Labour and Social Security (MLSS)		Government of Mexico: <i>SEDESOL</i>
Procurement process	Competitive procurement process	Competitive procurement process		Competitive procurement process
To whom evaluator is accountable	MCC	MLSS		<i>SEDESOL</i>

Table 1 (continued)

Features	Program		
	MCC Compact (El Salvador)	PATH (Jamaica)	Hábitat (Mexico)
Time period of evaluation	Five years: 2007-2012	Four years: 2003-2007	Two years: 2003-2004
Impact evaluation experience in the country	Limited	Very limited	Recent but rapidly growing
Implementing agency's impact evaluation experience	Very limited	Very limited	Limited

Notes: FOMILENIO = Fondo del Milenio; GDP = gross domestic program; MCC = Millennium Challenge Corporation; MLSS = Ministry of Labour and Social Security; PATH = Program for the Advancement Through Health and Education; SEDESOL = Secretaría de Desarrollo Social.
1/Information on the FOMILENIO intervention is limited to activities being evaluated by the authors. Activities of the FOMILENIO intervention that will not be evaluated by the authors (and are thus not mentioned above) include the Community Development activity within the Human Development Project and the Connectivity Project. 2/Expressed in 2008 U.S. dollars, adjusted using the Gross Domestic Product (GDP) Deflator inflation index. In 2006 U.S. dollars, the funding level for all MCC activities in El Salvador was \$461 million. 3/Expressed in 2008 U.S. dollars, adjusted using the Gross Domestic Product (GDP) Deflator inflation index. In 2006 U.S. dollars, the funding level for production and business services was \$57 million, and the funding level for investment support was \$21 million. 4/Expressed in 2008 U.S. dollars, adjusted using the Gross Domestic Product (GDP) Deflator inflation index. In 2006 U.S. dollars, the funding level for formal technical education was \$19.8 million, and the funding level for technical assistance was \$2.9 million. 5/Expressed in 2008 U.S. dollars, adjusted using the Gross Domestic Product (GDP) Deflator inflation index. In 2001 U.S. dollars, the funding level for PATH implementation from 2001-2005 was \$77.5 million: the loan from the World Bank was \$40 million and funding from the Government of Jamaica was \$37.5 million. 6/Expressed in 2008 U.S. dollars, adjusted using the Gross Domestic Product (GDP) Deflator inflation index. In 2005 U.S. dollars, the funding level for Hábitat from 2003 to 2006 was \$1.24 billion.

Sources: Levy and Ohls 2007; Campuzano and Levy 2007; Millennium Challenge Compact between the United States of America acting through the Millennium Challenge Corporation and the Government of the Republic of El Salvador 2006; Parrilat 2008, Ministry of Labor and Social Security 2008.

Table 2
 Barriers for Rigorous Impact Evaluation of Three Social Programs in
 El Salvador, Jamaica, and Mexico

Features	Program		
	MCC Compact (El Salvador)	PATH (Jamaica)	Hábitat (Mexico)
Pilot program intervention	Yes, for some components	Yes, in one geographic area (parish)	No
Pilot evaluated	No	No	No
Rigorous evaluation design required	Yes, MCC mandated that rigorous designs be used for impact assessments of major project components whenever feasible	Yes, strong support from a group within the government and from the World Bank	Unclear
Impact evaluation design used	Productive Development Project: experimental and/or quasi-experimental; possible experimental design—randomized phase-in or outreach; possible quasi-experimental design—comparison group of nonbeneficiary communities created using statistical matching Human Development Project, technical high school component: quasi-experimental comparison group of nonparticipating schools created using propensity score matching	Quasi-experimental: comparison group of ineligible families selected using regression discontinuity	Quasi-experimental: comparison group of nonbeneficiary census tracts selected using propensity score matching

Table 2 (continued)

		Program	
Features	MCC Compact (El Salvador)	PATH (Jamaica)	Hábitat (Mexico)
	Human Development Project, scholarship component: experimental, in which the probability of assignment to the intervention group varies according to merit and/or need		
Process by which design was selected	MCC listed design options in request for proposals; final designs were selected through negotiation among MCC, <i>FOMILENIO</i> , and the independent evaluator	Negotiation between the independent evaluator and MLSS	The independent evaluator proposed the matched-comparison group design and <i>SEDESOL</i> agreed to the design
Reasons for departing from an experimental design (if any)	Concerns about equity and fairness Concerns about disruption to program operations Use of randomization could lead to lower numbers of beneficiaries being reached, causing the government of El Salvador to fail to comply with a Compact provision with international treaty status <i>FOMILENIO</i> staff heard of bad experiences in similar programs in other countries	Random assignment judged to be operationally infeasible by MLSS	Evaluation began after program started
Timing of evaluation design	Concurrent with intervention design	Shortly after the intervention began	After the intervention began

Table 2 (continued)

Features	Program		
	MCC Compact (El Salvador)	PATH (Jamaica)	Hábitat (Mexico)
Design for primary data collection	<p>Productive Development Project: household surveys</p> <p>Human Development Project, technical high school component: school and student surveys</p> <p>Human Development Project, scholarship component: student surveys</p>	(1) household surveys of participant and comparison group households, (2) qualitative interviews with PATH and school and health care staff, and (3) focus group discussions with PATH clients	Not applicable
Entity responsible for data collection	Public institute and private contractor	Public institute	Public institute
Data collection was independent of the implementing organization	Productive Development Project: yes, tentatively	Yes	Yes
Potential to collect primary data	Human Development Project, technical high school component: limited independence	High, but there were issues with organization of data collection	None
Secondary data available	Human Development Project, scholarship component: unclear	MIS	Population census
	Management Information System (MIS)	Survey of Living Conditions, conducted by the Statistical Institute of Jamaica	
	Household surveys		
	Population, economic, and agricultural census		

Table 2 (continued)

Features	Program		
	MCC Compact (El Salvador)	PATH (Jamaica)	Hábitat (Mexico)
Adequacy of secondary data	Good	Good	Limited
Limitations of secondary data	Some surveys, including the <i>Censo Económico</i> , are nonrepresentative of El Salvador's Northern Zone	Living Conditions Survey was initially designed to oversample poor households, but the data collection agency did not implement this design	As data were limited to the 2000 Census and the 2005 <i>Conteo de Población y Vivienda</i> , only three infrastructure components could be considered and only the first two years of program implementation could be evaluated
Key outcomes	Productive Development Project: practices, employment, investment, productivity, and household and business income Human Development Project, technical high school component: retention and graduation rates, employment, and household income Human Development Project, scholarship component: retention and graduation rates, employment, and wages	School attendance and preventive health care use	Access to drinking water, sewage and drainage, and electricity

Note: FOMILENIO = Fondo del Milenio; MCC = Millennium Challenge Corporation; MLSS = Ministry of Labour and Social Security; PATH = Program of Advancement Through Health and Education; SEDESOL = Secretaría de Desarrollo Social.
Source: Levy and Ohls 2007, Campuzano and Levy 2007, Millennium Challenge Compact between the United States of America acting through the Millennium Challenge Corporation and the Government of the Republic of El Salvador 2006.

References

Attanasio, Orazio, Adriana Kugler, and Costas Meghir. “Training disadvantaged youth in Latin America: evidence from a randomized trial.” Working Paper No. 13931. Cambridge, MA: National Bureau of Economic Research, 2008.

Campuzano, Larissa, and Dan Levy. “The Effects of Hábitat on Basic Infrastructure.” Princeton, NJ: Mathematica Policy Research, Inc., December 2007.

Card, David, Pablo Ibararán, Ferdinando Regalia, David Rosas, and Yuri Soares. “The Labor Market Impacts of Youth Training in the Dominican Republic: Evidence from a Randomized Evaluation.” Working Paper No. 12883. Cambridge, MA: National Bureau of Economic Research, 2007.

CONEVAL. Consejo Nacional de Evaluación de la Política de Desarrollo Social. Available online at: http://www.coneval.gob.mx/coneval2/htmls/quienes_somos/HomeQuienesSomos.jsp?categorias=QUIENES,QUIENES-antecedentes (consulted on January 8, 2009).

Duflo, Esther, Rachel Glennerster, and Michael Kremer. “Using Randomization in Development Economics Research: A Toolkit.” Center for International Development Working Paper No. 138. Cambridge, MA: Center for International Development at Harvard University, December 2006.

Estados Unidos Mexicanos. “Lineamientos Generales para la Evaluación de Programas Federales de la Administración Pública Federal.” Mexico City: Diario Oficial, March 30, 2007.

Levine, Ruth E. “Multilateral Development Banks: Promoting Effectiveness and Fighting Corruption.” Testimony to the Senate Finance Committee on Foreign Relations. Washington, D.C.: Center for Global Development, March 28, 2006.

Levy, Dan, and James Ohls. “Evaluation of Jamaica’s PATH Program: Final Report.” Princeton, NJ: Mathematica Policy Research, Inc., March 2007.

Levy, Santiago. *Progress Against Poverty. Sustaining Mexico’s PROGRESA-OPORTUNIDADES Program.* Washington, D.C.: Brookings Institution Press, 2006.

Millennium Challenge Corporation. Millennium Challenge Corporation. Available online at: <http://www.mca.gov/countries/elsalvador/index.php> (consulted on December 3, 2008).

Ministry of Labor and Social Security. “Project Information Document Appraisal Stage.” Kingston, May 2008.

Parrillat, Gerardo Franco. “Migración y Desarrollo: Hacia Políticas Públicas Innovadoras en México: Programas de SEDESOL.” Mexico City: Secretaría de Desarrollo Social, February, 2008.

Phillips-Mandaville, Alicia. “MCC’s Approach to Country Ownership.” Millennium Challenge Corporation Working Paper. Washington, D.C.: Millennium Challenge Corporation, September 2008.

Pritchett, Lant. “It Pays to be Ignorant: A Simple Political Economy of Rigorous Program Evaluation.” *The Journal of Policy Reform* vol. 5, no. 4 (2002): 251-269.

Ravallion, Martin. “Should the Randomistas Rule?” *Economists’ Voice*, February, 2009. Available online at: www.bepress.com/ev (consulted on March 11, 2009).

Savedoff, William D., and Ruth Levine. “The Evaluation Agenda.” In *Rescuing the World Bank. Selected Essays.* Washington, D.C.: Center for Global Development, 2006.

Savedoff, William D., Ruth Levine, and Nancy Birdsall. “When Will We Ever Learn? Recommendations to Improve Social Development through Enhanced Impact Evaluation.” Washington, D.C.: Center for Global Development, September 15, 2005.

Treviño, Ernesto. “Diseños Experimentales: Evaluación de Programas de Transferencias Monetarias Condicionadas.” Paper presented at the seminar “La Importancia de Evaluar Políticas Educativas.” Santiago, Chile: Universidad Diego Portales, October 2008.

Un Buen Comienzo. Un Buen Comienzo: Educación Inicial de Calidad. Available online at: <http://www.unbuencomienzo.cl/> (consulted on December 3, 2008).

Wiebe, Franck S. “Aid Effectiveness: Putting Results at the Forefront.” Millennium Challenge Corporation Working Paper. Washington, D.C.: Millennium Challenge Corporation, October 2008.

