



Cameron Breslin, Alexandra Orsola-Vidal, Temina Madon

An Innovative Three-pronged Approach to Institutionalise Impact Evaluation Within Multilaterals

We live in a world in which governments, multilaterals, and non-governmental organisations are increasingly committed to integrating impact evaluation into development programs. Never has there been stronger demand for evidence within the development sector. The full array of stakeholders—from donor agencies to beneficiaries in-country—is demanding to know whether programs and policies are meeting their intended objectives. These stakeholders seek empirical evidence that agencies are investing finite resources in the most efficient and effective approaches. Without rigorous evaluation, decision makers risk making poor choices about which programs to fund and which to terminate.

Despite the need and demand for rigorous evidence, impact evaluation has yet to become fully institutionalized within multilateral development agencies.¹ Recognizing this, the Center for Effective Global Action (CEGA), headquartered at the University of California, Berkeley, has created a unique program to build capacity and promote the mainstreaming of impact evaluation within multilaterals. CEGA, a research network that designs and tests solutions for the problems of poverty using rigorous evaluation techniques, behavioural experiments, and tools from data science, counts 60 faculty researchers across 9 U.S. universities, working in partnership with researchers in more than 20 developing countries.

The program, designed and piloted by CEGA, aims to institutionalize impact evaluation within agencies and their government counterparts. It generates buy-in among stakeholders while building their capacity. It also creates agency and government access to a network of top academic researchers for on-going consultation and hands-on mentorship. The program has been piloted with the Inter-American Development Bank (IDB), demonstrating a scalable approach that has the potential to benefit other large, multilateral donor agencies.

Designing an Intervention for Multilaterals

In 2012, CEGA established the *Berkeley-IDB Impact Evaluation Collaborative* (BIC) after a year-long collaborative design process involving UC Berkeley faculty and IDB staff. The design drew on the experience of UC Berkeley economist Paul Gertler, former Chief Economist for Human Development at the World Bank, and of IDB economist Sebastian Martinez, also formerly of the World Bank.

During their tenure at the World Bank, Gertler and Martinez had been confronted by the many challenges of implementing impact evaluations within a multilateral agency. Foremost among these was the need to coordinate and synchronize the timelines, priorities, and assets of various stakeholders. Ultimately, weaving rigorous evaluation into core operations required a multi-pronged approach. Gertler and Martinez identified at least three key components: 1) creating durable mechanisms for linking academic expertise with operations and government professionals; 2) generating buy-in for impact evaluation among key actors, and 3) building the capacity of agency and government stakeholders to guide and directly participate in evaluations.

Based on these insights, BIC was designed to incorporate multiple activities:

- *Project-based executive education*: BIC is built around an annual, two-week executive education course for IDB staff and government counterparts, held in person but off-site, to minimize distraction. Participants are accepted as teams, and come to the course prepared to work on

1 <http://www.cgdev.org/publication/when-will-we-ever-learn-improving-lives-through-impact-evaluation>

the evaluation of a specific IDB-funded project. Projects are vetted internally by IDB, based on the feasibility of rigorous evaluation. Each project team includes multiple members to build buy-in and momentum, incorporate diverse perspectives and priorities, and maximize the likelihood of an evaluation moving forward. As part of the two-week course, participants are introduced to impact evaluation methods via lectures, group work, and case-based learning from top experts in the field. The curriculum focuses on research design, statistical methods, survey management, and dissemination of results to policymakers. At multiple periods throughout the course, participants work directly on their IDB-funded projects, each of which has the potential to translate into a multi-year program evaluation with long-term impact. These sessions are mentored by experts and make it possible to immediately apply methods to real projects. Teams thus not only gain new skills but also develop an evaluation action plan co-developed by key stakeholders.

- *Continued online learning:* After the two-week executive course, participants are offered the opportunity to extend their learning for an entire calendar year through an online distance-learning course developed by CEGA and hosted by the edX partnership. The online course is considerably more technical than the on-site workshop. It focuses on the analysis of impact evaluation data

using Stata and other statistical software packages. Each online module builds on the methodology lectures and case studies introduced during the two-week on-site training but also includes video content and problem sets that provide hands-on experience in statistical modelling, power calculations, data management, and analysis. The online course also incorporates topics beyond the scope of the two-week course, such as survey management, research transparency, and the protection of human subjects. Each module involves the analysis or review of data from real-world evaluations, further enabling participants to implement their own program evaluations.

- *On-going mentorship from research experts:* At the conclusion of the two-week on-site training, any team with a viable project design is invited to receive one year of mentored support from faculty experts in the CEGA network. Mentors provide both remote and in-person guidance throughout the evaluation design phase. Faculty member time can be used to solidify the team's skills, expand or revise the evaluation design, identify funding for survey data collection, and develop longer-term partnerships. At the close of the year, representatives from each project team are invited to UC Berkeley to present and refine their evaluation designs while receiving training in research dissemination, including strategies to influence high-level decision-making and generate public buy-in for evidence-based programs.



Paul Gertler, Li Ka Shing Professor of Economics at Haas School of Business and the School of Public Health at University of California, Berkeley, lectures to BIC participants during the second year of the conference (July 2014).

Program Implementation

BIC is currently in its third year of implementation. The first executive education course began in July 2013 and the second was held in July 2014. In total, 115 participants have attended the intensive two-week training. The program has sought to achieve an even balance between IDB staff and government counterparts; 47 per cent of participants represent the government arm responsible for implementation and/or evaluation of the team's development program. Over the two years, participants have represented 36 IDB-funded projects from 13 countries across Latin America and the Caribbean.

During the first year, BIC focused on infrastructure, including projects on agriculture and climate change, water and sanitation, energy, housing, and transportation. As such, the faculty experts providing lectures and mentoring were drawn from a range of disciplines, including agricultural and resource economics, energy economics, engineering, and business. In addition to providing instruction during the two-week executive education, the faculty members also provided remote mentorship to project teams in the year following the initial executive education. During its second year, BIC focused on early childhood development, health, and education. With team projects in these sectors, the composition of faculty expertise was tailored to include public health and education researchers, economists, and political scientists. As they did during the first year, the experts played roles in both the instruction and remote mentorship of participants.

The workshop itself is designed to integrate executive education (typically delivered through lectures and case studies) with hands-on "learning by doing." To this end, four different types of sessions are woven into the cohesive two-week program:

1. *Lecture and panel sessions* outline the different methods and tools for impact assessment, including both experimental and quasi-experimental methodologies. These sessions are divided into a technical track and a policy track (see below).
2. *Field methods sessions* build familiarity with the necessary steps in implementing an impact evaluation, including budgeting, survey design, power calculations, and quality assurance.
3. *Sector-specific breakout sessions* are essential seminar-style presentations by leading academics to expose participants to the latest evidence and evaluation research in each sector or focus area.
4. *Mentored group work sessions* create opportunities for IDB project teams to work directly with faculty members to design impact evaluations for active Bank projects.

The first week of the workshop provides an overview of basic statistics, evaluation theory and methods, and advanced topics in impact evaluation. The key concepts are taught primarily through case study discussions, lectures, and panels. The second week focuses on the process of implementing rigorous evaluations, with group exercises and discussions interspersed with lecture and panel sessions, and aims to cement learning through practice. Both weeks feature afternoon "industry seminars" with academics from different disciplines presenting on the latest research advances in key sectors. These include engineering advances as well as new tools for measuring development outcomes (such as sensors, mobile devices, and methods from data science).

Over the course of the training, six sessions are reserved for mentored teamwork, led by faculty members and Ph.D. student researchers. This is the time when participants apply evaluation methods to their IDB-financed projects. On any given day, the teamwork is tied directly to the techniques taught during the morning's plenary sessions. Group work progresses step-wise, from outlining a program's theory of change to defining hypotheses and identifying counterfactuals. As a final output, each team delivers a presentation outlining the proposed evaluation design, required resources, and timeline. On the last day of the

two-week training, each team presents its design for peer review and feedback. BIC instructors and staff evaluate the presentations for viability, using four criteria:

1. **Significance:** Does the study make a significant contribution toward advancing IDB's knowledge of international development in Latin America and the Caribbean?
2. **Innovation:** Does the study answer new questions, or introduce novel methods, measures, or interventions?
3. **Viability:** Is the research design appropriate and well-articulated? Will the study be able to answer proposed questions? Are the indicators and sample size estimates appropriate, given the outcomes to be measured? Does the proposal address logistical or political obstacles that might threaten the completion of the study?
4. **Translation:** Is the program relevant to the IDB and to projects in other IDB countries? Can it be applied to existing programs or initiatives in the region?

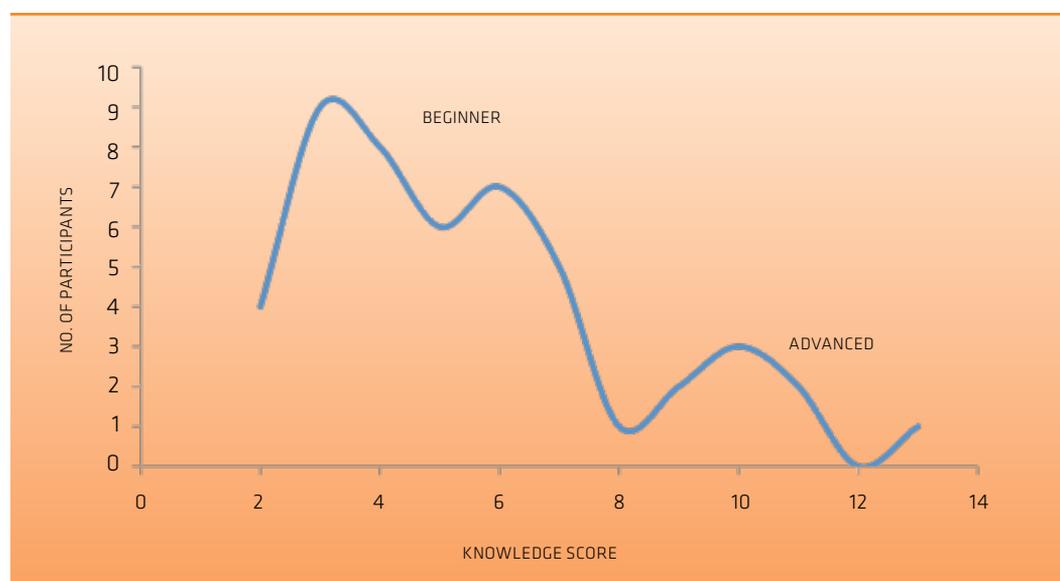
Projects are selected for follow-up support based on expressed demand by the project team and the BIC faculty's assessment of viability. To date, 12 projects have been selected over two years to receive continued support from faculty.

Results and Learning

An important component of BIC has been its commitment to continuous evaluation and data-driven redesign. Several instruments are used to collect data about the program, including participant applications, pre- and post-course exams, short daily surveys of course participants, and evaluation concept notes. In addition, administrative data are collected throughout the program's implementation—including terms of reference for faculty mentoring grant proposals for specific evaluations, and routine reports. Analysis of these data has informed important program modifications.

For example, during the first year of the workshop, it was observed that the participants' impact evaluation knowledge and technical skills were heterogeneous (see Figure 1). To better meet the needs of the diverse participant pool, two tracks were created for the second year of instruction: a

Figure 1 Distribution of Participants' Prior Knowledge



Note: The distribution of knowledge scores for participants in Year 1 shows a bimodal distribution, suggesting separate populations of beginner and advanced individuals. Data are from a pre-course exam administered to all participants.

Technical Track and a Policy Track. The two-track system ensures that participants are exposed to the appropriate level of content while still learning to speak the same impact evaluation “language” by the end of the course.

The Technical Track, taught in English, has been designed to meet the needs of participants who have previous experience in impact evaluation in addition to an academic background in econometrics or in statistics. Technical Track participants are likely to be applying evaluation techniques in their day-to-day work already, or intend to do so. They are expected to be familiar with both experimental and quasi-experimental field methods. Instruction is delivered in a series of lectures in applied econometrics covering several advanced methods and statistical models in detail. This track also explores new techniques for applying methods in the field, including approaches that minimize potential risks to study design.

The Policy Track, taught in Spanish, is designed for development professionals who have not been

exposed to impact evaluation. Participants may be interested in understanding how evaluation is used to improve the quality of programs, or how evidence can drive decision-making among policy-makers and constituents. The Policy Track requires no background in statistics or econometrics, and sessions are taught through an intuitive approach with minimal use of equations or statistical models. Participants learn the value of impact evaluation methods; they also explore how and why different methods meet the needs of different programs. They learn how to identify evaluations that credibly demonstrate a program’s casual impact. This track provides an overview of experimental and non-experimental impact evaluation methodologies, and explores the intuitions and assumptions and strengths and weaknesses behind each method.

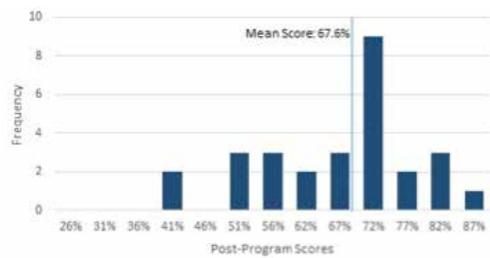
As shown in Figure 2, participants in both the Policy and Technical Tracks have demonstrated increases in knowledge by the end of the two-week executive education. Policy Track participants showed a 31 per cent increase on average between their pre-program assessment scores and

Figure 2 Pre- and Post-Workshop Assessment Scores

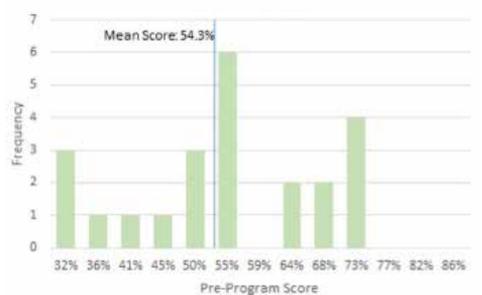
DISTRIBUTION OF PRE-PROGRAM SCORES - POLICY TRACK



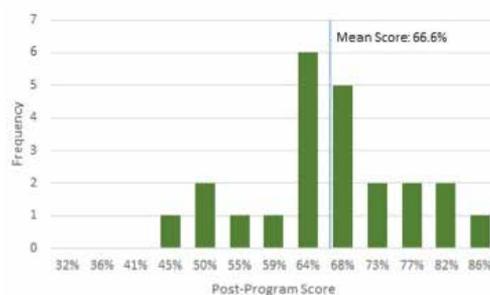
DISTRIBUTION OF POST-PROGRAM SCORES - POLICY TRACK



DISTRIBUTION OF PRE-PROGRAM SCORES - TECHNICAL TRACK



DISTRIBUTION OF POST-PROGRAM SCORES - TECHNICAL TRACK



their post-program scores. Scores of those in the Technical Track rose by 23 per cent on average. Changes in scores are statistically significant for both tracks ($p < 0.01$). Different assessment tools were used for the two tracks. While mean scores are low in both groups (i.e. the assessment exam was difficult for participants), there is evidence that the program increased their knowledge of impact evaluation.

Importantly, participants also expressed strong satisfaction with the program, assigning an average score of 6 of 7. They found the workshop to be highly relevant to their productivity and to the quality of their work. Seventy per cent of participants responded that concepts and skills learned in the executive education were immediately applicable to their jobs. This suggests that the program was also successful in generating buy-in for impact evaluation, which is an essential step in institutionalizing the approach.

BIC has also yielded results in the implementation of impact evaluations. Through participation in the program, IDB evaluation teams have leveraged over USD 1.5 million in funding for impact evaluations now in the field. At least 7 project teams continue to draw on university expertise through active collaboration on evaluations, suggesting that BIC has been able to broker enduring relationships between academia and implementing agencies.



Erin Murphy-Graham (Professor at UC Berkeley's School of Education) and Michelle Perez (IDB) mentor an IDB team evaluating the effectiveness of an early childhood nutrition program during the 2nd year of the BIC program (July 2014).

These studies are evaluating programs to reduce poverty through conditional cash transfers, health and nutrition interventions, improved early and primary education, labour market opportunities, and gender violence prevention.

A strong path forward

Building on experience within the World Bank, CEGA and the IDB have developed a new approach to capacity building and to the institutionalization of impact evaluation within multilateral agencies. The World Bank Independent Evaluation Group² has suggested that progress toward the institutionalisation of evaluation can be demonstrated through:

1. **Country-led initiatives:** Evaluations are country-led and managed by a central government or a major sectoral agency.
2. **Strong “buy-in” from key stakeholders:** There is strong acceptance among the agencies being evaluated, parliament, policymakers, budget planners, and the public—and strong support of a powerful central government agency (usually finance or planning) that manages the program's implementation. Although the system needs “champions,” particularly in its early stages, it is essential that it be depoliticised so that it is not seriously affected by national elections or changes in the administration.
3. **Evaluation capacity development:** Capacity must be developed to commission, design, conduct, manage, and use IEs. There is both a strong demand for the evaluations and an adequate supply of technical expertise and the organizational capacity to conduct evaluations and analyse data. Data-collection and analysis capacity has also been developed in planning agencies so that data sets, such as household

² See “Institutionalizing Impact Evaluation within the Framework of a Monitoring and Evaluation System” http://econ.worldbank.org/external/default/main?pagePK=64165259&theSitePK=469372&piPK=64165421&menuPK=64166093&entit yID=000333037_20091217225639

income and expenditure or demographic and health surveys, are available for use as baseline data or for selecting control/comparison groups for new studies.

The training model tested through BIC aims to achieve these three targets. Project teams intentionally include government implementers and multilateral staff, in order to make evaluations more “country-led.” These teams also generate strong buy-in across key stakeholders. The program seeks to build the capacity of all stakeholders involved and teaches to the appropriate level for each stakeholder, dividing participants into Policy and Technical tracks. This ensures that all parties increase their skills while sharing a common vocabulary. The program has been shown to increase the participants’ knowledge of impact evaluation methodologies, as their technical skills. It has also resulted in the implementation of multiple on-going evaluations.

CEGA’s multi-pronged approach to the institutionalisation of impact evaluation within multilateral

donor agencies has only been implemented with the IDB thus far, although it has the potential to succeed within other multilateral donor agencies. However, promoting impact evaluation in new contexts and institutions will require us to tailor the program to relevant cultures, norms, and incentive structures.



Alexandra Orsola-Vidal (CEGA) mentors an IDB team evaluating the effectiveness of a program intended to improve road safety in Colombia during the first year of the BIC program (July 2013).

PROFILE OF THE AUTHORS

Cameron Breslin supports CEGA’s training and capacity building initiatives as Program Associate, managing the Berkeley-IDB Impact Evaluation Collaborative. He is also Research Assistant to Professor Paul Gertler. Cameron has extensive experience in training and capacity building, particularly for impact evaluation, having previously led staff capacity building activities at Innovations for Poverty Action. He has conducted fieldwork measuring the effects of microfinance products in Bangladesh and financial literacy education programs in Uganda. He holds a BA in Comparative Literature from Hamilton College.

Alexandra Orsola-Vidal is the Global Networks Director at CEGA. She directs all training and capacity building initiatives and leads CEGA’s efforts to create a network of universities, development agencies and local research institutions that promote rigorous program evaluation. Since 2006, she has consulted for the World Bank, designing and implementing randomized controlled trials of large-scale development programs globally. She has extensive expertise facilitating impact evaluation workshops for local counterparts and bilateral agencies. Previously, she worked for international and local NGOs such as BRAC and Engineers Without Borders. Alex holds an MSc in Development Studies from the London School of Economics and a BA in Economics from the Universitat Autònoma de Barcelona.

Temina Madon is the Executive Director of CEGA and provides leadership in the Center’s scientific development, partnerships, and outreach. She has worked as science policy advisor for the National Institutes of Health Fogarty International Center, where she focused on enhancing research capacity in developing countries. She has also served as Science and Technology Policy Fellow for the U.S. Senate Committee on Health, Education, Labor and Pensions, managing an extensive portfolio of global health policy issues. She holds a Ph.D. from U.C. Berkeley and an S.B. from MIT.