

Development Impact Evaluation Initiative: A World Bank-Wide Strategic Approach to Enhance Developmental Effectiveness

Arianna Legovini

Head, Development Impact Evaluation Initiative

June 29, 2010

To contact the author, please email alegovini@worldbank.org. Jishnu Das, Maria Jones, Priscila Malaguti, Mulalo Muthige, Joao Luis Soares and Jee-Peng Tan contributed examples and perspectives. Comments from Asli Demirguc-Kunt, David Evans, Florence Kondylis, Nandini Krishnan, Ariel Fiszbein, David McKenzie, Jishnu Das, Victor Orozco and Mohammad Zia Qureshi are gratefully acknowledged and so are the discussions with Abhijit Banerjee. Further comments were discussed with members of the DIME Steering Group including Ana Revenga, Adam Wagstaff, Aysegul Akin-Karasapan, Emmanuel Skoufias, James Parks, Laura Rawlings, Marianne Fay, Martin Ravallion, Nadeem Mohammad, and Shahrokh Fardoust. The support of Justin Lin is gratefully acknowledged as are the many contributors to the OVP discussion. The author would like to thank Mattea Stein, Isabel Beltran, Fernanda Luchine, Ozan Sevimli and Johan Mistiaen for help with the updating of the DIME database and the estimation of costs. Further, the author would like to thank all the Bank TTLs and managers, internal and external researchers and the whole DIME team for working so hard to make DIME happen. Special thanks to Graeme Wheeler for his leadership and support.

Abbreviations

3IE	International Initiative for Impact Evaluation
AAA	Analytical and Advisory Services
AADAPT	Agricultural Adaptations impact evaluation program
ACT	Artemisinin Combination Therapy
ADP-SP	Agricultural Development Program Support Project (Malawi)
AFR	Africa Region
AGEMAD	Amélioration de la Gestion de l'Education à Madagascar
AIDS	Acquired Immunodeficiency Syndrome
AIM	Africa Impact Evaluation Initiative
AIM-AIDS	Africa Impact Evaluation of HIV/AIDS program
ALMP	Active Labor Market Program
APEIE	Africa Program of Education Impact Evaluation
ARD	Agriculture and Rural Development
ART	Anti-retroviral Treatment
BB	Bank Budget
BNPP	Bank Netherlands Partnership Program
BPRP	Belgian Poverty Reduction Partnership
CCT	Conditional Cash Transfer
CDD	Community Driven Development
CEGA	Center of Evaluation for Global Action
CLEAR	Regional Centers for Learning on Evaluation and Results
CMU	Country Management Unit
CSPRO	Census and Survey Processing System
DEC	Development Economics
DECDG	Development Economics Data Group
DECOS	Development Economics Operations and Strategy Group
DECRG	Development Economics Research Group
DFID	Department for International Development (UK)
DIEFS	Developmental Impact Evaluation in Fragile States
DIME	Development Impact Evaluation Initiative
DIME-FPD	Development Impact Evaluation in Finance and Private Sector
EAP	East Asia and Pacific Region
ECA	Europe and Central Asia Region
ECD	Early Childhood Development
EGAP	Experiments in Governance and Politics
EPAG	Economic Empowerment of Adolescent Girls (Liberia)
EPDF	Education Program Development Fund
EFI-FTI	Education-For-All Fast-Track Initiative
ETC	Extended Term Consultant
FPD	Finance and Private Sector
FT	Full Time
FY	Financial Year
GAP	Gender Action Plan
HD	Human Development
HDN	Human Development Network
HIV	Human Immunodeficiency Virus
HRBF	Health Results-Based Financing
IBRD	International Bank for Reconstruction and Development
IDA	International Development Association
IDB	Inter-American Development Bank
IDF	Institutional Development Fund
IE	Impact Evaluation
IEG	Independent Evaluation Group
IL	Investment Lending
ISR	Implementation Status and Results Report
ISUP	Informal Settlements Upgrading Program (South Africa)

IT	Information Technology
IV	Instrumental Variables
JPAL	Abdul Latif Jameel Poverty Action Lab
KCP	Knowledge for Change Program
KP	Knowledge Management Product
LAC	Latin America and Caribbean Region
LLINs	Long-Lasting Insecticide-Treated Nets
LPRP	Luxembourg Poverty Reduction Partnership
LSE	London School of Economics and Political Science
LSMS	Living Standards Measurement Surveys
M&E	Monitoring & Evaluation
MCC	Millennium Challenge Corporation
MD	Managing Director
MDGs	Millennium Development Goals
MENA	Middle East and North Africa Region
MIC	Middle Income Countries
MIEP	Malaria Impact Evaluation Program
MIT	Massachusetts Institute of Technology
MoAFS	Ministry of Agriculture & Food Security (Malawi)
NDHS	National Department of Human Settlements (South Africa)
NREG	National Rural Employment Guarantee (India)
NUSAF	Northern Uganda Social Action Fund
OPCS	Operations Policy and Country Services
OVP	Operational Vice-Presidents
PREM	Poverty Reduction and Economic Policy Network
PRMGE PREM	Gender and Development
PROGRESA	Programa de Educación, Salud y Alimentación
RDD	Regression Discontinuity Design
RF	Research Services
RSB	Research Support Budget
RSBY	Rashtriya Swasthya Bima Yojana (India)
SAP	Systems, Applications and Products in Data Processing
SAR	South Asia Region
SDN	Sustainable Development Network
SIEF	Spanish Impact Evaluation Fund
STC	Short Term Consultant
TA	Technical Assistance
TF	Trust Fund
TFESSD	Trust Fund for Environment and Socially Sustainable Development
UC	University of California
UNAIDS/UBW	UNAIDS Unified Budget and Work plan
USAID	United States Agency for International Development
WB	World Bank
WDR	World Development Report
YDP	Youth Development Program
YE	Youth Employment

Table of contents

Arianna Legovini	i
Head, Development Impact Evaluation Initiative	i
Executive Summary	1
Introduction	5
DIME History	6
DIME Objectives	8
DIME Strategy	9
Step 1: A results-based product to improve operational quality	9
Step 2: Building capacity through joint evaluations	12
Step 3: From local learning to global learning	18
DIME Thematic Programs	20
DIME Methods	28
DIME Contribution to the Results Agenda	33
Data quality, availability and usage	33
M&E systems	34
Attribution	35
DIME Human and Financial Resources	37
Skills	37
Costs and financing	38
Conclusions	43
References	44
Annex A. DIME Governance Structure	47
Annex B. Description of Thematic Impact Evaluation Programs	49
I. Finance and Private Sector Program	49
II. Human Development Programs	50
III. Sustainable Development Programs	55
IV. Poverty Reduction and Economic Policy Programs	57
V. Multi Sector Programs	60
Annex C. Academic partners	62
Annex D. DIME Data Catalogue	63
Boxes	
Box 1. Education in Punjab, Pakistan	9
Box 2. Fine-tuning Brazil's Descomplicar	12
Box 3. "We can make a real difference in future policy developments": South Africa's National Department of Human Settlements	13
Box 4. Africa Program for Education Impact Evaluation (APEIE)	14

Box 5. “We won’t be passive passengers”: Malawi’s evaluation of peer farmers.....	15
Box 6. Building capacity for Madagascar’s improvement of school management.....	17
Box 7. Impact evaluation glossary.....	29
Box 8. Getting M&E right: India’s health insurance scheme.....	36
Box 9. The Spanish Impact Evaluation Fund (SIEF).....	50

Figures

Figure 1. Evolution of Bank IE, 2004 – 2010.....	7
Figure 2. IE distribution by region.....	7
Figure 3. IE distribution by network.....	7
Figure 4. DIME governance structure.....	8
Figure 5. Test nodes of decision tree to improve program effectiveness.....	10
Figure 6. Active Impact Evaluations.....	18
Figure 7. DIME thematic programs.....	21
Figure 8. Organization of a thematic impact evaluation program.....	22
Figure 9. IE in lending by region, FY 2010.....	23
Figure 10. IE in lending by network, FY 2010.....	23
Figure 11. Quality Assurance.....	24
Figure 12. Number of IEs started by fiscal year, 1976 - 2010.....	25
Figure 13. IE by thematic program.....	27
Figure 14. Randomized Methods.....	30
Figure 15. Use of multiple methods.....	31
Figure 16. Staff with impact evaluation skills.....	37
Figure 17. IE staff by level.....	38
Figure 18. IE staff by contract type.....	38
Figure 19. Bank-executed IE spending, FY 2000-2010.....	39
Figure 20. Spending by product type, FY 2006 - 2010.....	40

Tables

Table 1. Africa Program for Education Impact Evaluation budget (14 countries), FY 2007-2010.....	40
Table 2. Cross-country impact evaluation workshops.....	41
Table 3. Cost structure of impact evaluation products.....	42

Executive Summary

Background

Created in 2005 in the Bank's Chief Economist Office, DIME was re-launched in 2009 as a broad-based decentralized effort to mainstream the use of impact evaluation in the Bank. The effort is led by a high level steering group, coordinated by a secretariat based in DEC, and implemented by all networks and regions. DIME's objective is to improve the quality of the Bank's operations, strengthen country institutions for evidence-based policy making, and generate knowledge in 15 strategic development areas (World Bank 2009b).

Link to the results agenda

Impact evaluation falls within the larger context of the Bank's monitoring and evaluation (M&E) efforts and the results agenda (RA). The M&E effort aims at ensuring that all Bank's instruments count with a results framework and a data collection plan. M&E also supports governments and operations build monitoring systems. The results agenda systematizes the use of results Bank-wide by introducing the use of common indicators in Bank's operations, stimulating the adoption of a results-based culture, and shifting from input-based to output- and outcome-based reporting. By tracking change over time, M&E and RA provide a descriptive view of whether things are moving in the right direction.

Impact evaluation adds causal inference to the results agenda. Causal inference links the observed change in outcomes to specific policy actions. By measuring cause-effect relationships, IE helps guide whether an action should be maintained and the financing scaled up or down. To increase impact evaluation in Bank's operations, DIME makes available specialized teams and structures to provide them with just-in-time advice to incorporate international evidence in project designs, measure project results and improve impacts during implementation.

Impact evaluation currently covers 13 percent of the portfolio of active IBRD/IDA lending. Through its partnerships, DIME provides in-house technical assistance to the majority of these projects and clients throughout the project cycle to ensure high quality monitoring and impact evaluation. As a result, each project assisted includes a high-quality baseline/endline strategy for data collection to inform progress (to meet the monitoring objective and ascertain that things are moving in the right direction) and a robust analytical strategy to attribute the results to their interventions (to understand causal links and inform policy). The collaboration between research and operations ensures that data is of good quality, and that the baseline and endline are used for analysis and not only to calculate summary statistics. Moreover, the lending product is strengthened by a learning strategy that helps the project adapt to new, analytically sound evidence and secure greater effectiveness.

Institutional development focus

Working through the Bank's extensive network of relationships with government, DIME commits to work in full collaboration with country institutions for the medium term and throughout a full iterative process of learning. DIME specializes in evaluating programs at scale in the country's main institutional environment, and, in so doing, obtaining results

that are valid for policy decisions in the country-wide context. The collaborative process is used to transfer the technology to support evidence-based policy-making to country institutions through a learning-by-doing approach. By using impact evaluation both to measure and improve results, DIME is helping institutions use impact evaluation not simply as an accountability tool but also as a management tool to generate and use evidence for better policy.

Applicability to Bank lending products

Impact evaluation measures the effectiveness of specific interventions or programs and does not concern itself with the underlying financing mechanism. This is why impact evaluation is relevant not only to interventions that are financed through investment lending but also to those supported by development policy lending.

Links to the knowledge agenda

With 170 completed and 280 active studies in 72 countries to date, DIME is the largest initiative in the world designed to systematically learn from development experience on the basis of rigorous impact evaluation. To implement this effort, DIME works in collaboration with 80 World Bank researchers and with academics from more than 100 institutions. The top internal and external academic resources made available to policy makers on the ground create a powerful interface for building capacity and vibrant networks for generating ideas and knowledge. Working across 15 strategic knowledge areas identified through a Bank-wide consultative process, DIME generates and disseminates the lessons that will guide development effectiveness into the next decade.

DIME disseminates the results via cross-country workshops, seminars, conferences, network weeks, web and journal publications, and thematic policy reports. All IE results are published in the IE Working Paper Series and in the external IE database in a way that is accessible to all audiences. More importantly, during the DIME workshops, the results of completed evaluations are presented to the teams preparing new operations. The presentations are followed by project specific round tables that are facilitated by IE experts and that stimulate discussion around the evidence that is relevant to the design choices of those operations. The design of the average operation participating in DIME changes during the workshop. While figures supporting this claim are not currently available, DIME plans to track this indicator in the future. Further, presentations of IE results have been held during Network Weeks, and in the AIM, DIME and HDN seminar series. Finally, results from completed evaluations are used to inform proposed sector strategies through direct inputs or through DEC comments to the OVP discussions.

Policy research reports are another way of making IE results available to larger audiences. By the end of this calendar year, three policy research reports that summarize international IE evidence in CCT, local governance and educational accountability and two IEG reports summarizing IE lessons in nutrition and social safety nets will have been completed. They are one of the tools the Bank uses to help shift the development paradigm from prescription-based to evidence-based policy making.

Link with cost-benefit and developmental effectiveness

Impact evaluation estimates the benefit side of cost-benefit analysis. The growing availability of impact evaluation results is thus a unique and currently unexploited

opportunity for conducting ex-post cost-benefit analyses of Bank-financed interventions on a large scale. Availability of cost-benefit analysis for a large sample of Bank supported interventions across several sectors can be of great value to the development effectiveness agenda. It would enable development practitioners to compare the relative benefits of allocating resources to different interventions, potentially helping countries improve allocative efficiency and increasing the effectiveness of total financing.

Progress to date

Starting from a very low base in 2004, DIME's capacity to reach out to operations across the institution and disseminate the findings from completed and ongoing studies has grown. The initiative engages teams in a process of fine-tuning designs, developing learning strategies to generate evidence during implementation, and introducing the operational flexibility needed to incorporate new evidence over time. DIME has also started to reach outside the boundaries of the Bank's operations, with various countries and development institutions requesting assistance to develop impact evaluation structures. In addition, academics around the world work through DIME to lend their expertise and knowledge to policy-makers and national programs. Country institutions are the focus of DIME activity. Working now with about two hundred government agencies, DIME is helping them strengthen their planning and policy-making functions by providing them access to training, networks of practitioners and top quality analytical resources.

Governance structure

The Bank's Managing Director responsible for knowledge and networks established DIME's governance structure in December 2008. It transformed DIME into a Bank-wide decentralized program, coordinated by Development Economics (DEC) and spanning all regions and networks. The structure includes a DIME Steering Group (SG) responsible for setting Bank-wide priorities. The SG has chief economist and director level representation from networks and regions, reports to the Managing Director and the Chief Economist, and promotes the initiative in the networks and regions. The DIME Secretariat in DEC provides technical leadership to operationalize the SG decisions. The networks hold thematic responsibility and the regions are responsible for implementation. Analytical and data services are provided by the research and data groups in DEC.

Alignment with the Bank's learning priorities

Under this new structure, DIME is sponsoring 15 thematic impact evaluation programs agreed through a process of consultation and aligned with the Bank's learning priorities. They investigate the accountability structures that improve educational quality, the mechanisms for strengthening prevention in malaria and HIV, low-cost options for early childhood development, instruments for strengthening social protection, policies to lower constraints to private sector growth, reforms to improve governance, adaptations and mitigations to climate change in agriculture, water, energy and forestry, the economics of gender and mechanisms for state and peace building in fragile institutional contexts. Of the 15 programs:

- Eight were active before 2009 (education, health, HIV, malaria, CCTs, active labor markets, local development, and access to infrastructure);

- Four were launched during fiscal year 2009-10 (early childhood development, finance and private sector, institutional reform and agricultural adaptations) together with two cross cutting themes (gender and fragile states); and
- Three more are waiting to be launched (adaptations in forestry, water resource management, and energy mitigation).

Expansion of sector coverage increased demands from project teams and government agencies leading to a 46 percent growth in the DIME's portfolio between 2008 and 2010. While a large portion of the response was organized by the DIME Secretariat, most of the products were quickly incorporated in regional (61%) and network (17%) work plans, with DEC retaining only a minority of the products (22%).

Adaptation

While not discarding impact evaluation for accountability purposes (*what works*), DIME has embraced the use of IE to understand *how to* make policy work better. DIME's focus on a participatory process to empower program managers to set their own agendas is paying off. Government agencies looking for results are enthusiastic. However, in the transition from prescription to discovery, clients and World Bank teams may face adverse incentives and unclear rewards for more rigorous measurement of results. Uncertainty in financing at the time when clients commit to doing impact evaluation may also discourage adoption.

Cost-effectiveness

The cost of impact evaluation products is small relative to the cost of the interventions they evaluate. During fiscal years 2006 to 2010, the Bank's actual internal spending per impact evaluation product was \$54,000, leveraging an average of \$123,000 in grants and a quarter million dollars in government funding.¹ Total Bank-executed spending on IE represents, on average, 0.2 percent of project costs. Total IE costs (Bank-executed plus government contributions) in programs coordinated by the DIME Secretariat, for which figures exist, are about \$530,000 or 0.6 percent of project cost. The government contributions pay for the data which are used for both monitoring and evaluation.

Significant resources are used for capacity development. Through the cross-country workshops conducted by the Africa Impact Evaluation (AIM) Initiative, the DIME Secretariat and HDN from fiscal 2005 to fiscal 2010, 3,100 people from 450 teams were trained in impact evaluation and exposed to international evidence at a cost of \$1,400 per person or \$9,900 per team. 144 teams (or 86 percent) are implementing their impact evaluation out of the 167 that participated in the AIM and DIME workshops, for which data exist. At this efficiency rate, costs are \$11,500 per impact evaluation design that gets implemented.

¹ Source: SAP actuals.

Introduction

In the last decade, the demand for developmental effectiveness has become the norm. The standards for demonstrating impact have been raised. The development community is assuming its responsibility for learning systematically from experience using methods with high scientific rigor. The increased willingness of donors to finance, of governments to implement, and of academia to support rigorous impact evaluation efforts attest to it. The use and applicability of impact evaluation methods has been expanding rapidly because of the promise they hold: to significantly shift our knowledge of the *what* and the *how* of economic development.

Measuring causal links is of utmost significance for policy decisions. Spurious correlations may result in wrong policies.² When governments react to an economic recession by toughening immigration policy, for example, the slow-down in migration flows is often attributed to the enforcement efforts whereas it is the slowed economic conditions that reduce the pull on migrants. This type of policy error often goes unnoticed. Identifying the correct causal relationship can guide policies in the right direction.

Done well, impact evaluations can deliver precise estimates of the cause-effect relationship between policy action and outcomes by comparing predefined treatment and control groups before and after a policy intervention. This is referred to as causal inference or counterfactual analysis. Establishing the direction and magnitude of causal relationships can guide policy makers and donors regarding which policy should be used to achieve the desired objective, identify the policy's cost-effectiveness, and justify a response or a scale-up. When used to compare operational alternatives, impact evaluations can guide and improve the effectiveness of policy interventions. This requires that impact evaluations be prospective; in other words, that they be designed before an intervention takes place in order to measure, during implementation, the causal links between an intervention and changes in short-, medium-, or longer-term outcomes.

The Bank is implementing impact evaluation as a multiyear exercise, closely connected to program implementation. The aim is to turn impact evaluation into a package of products designed to inform policy making at multiple entry points during the Bank's project cycle and the country's policy cycle. When fully integrated into policy makers' and program managers' work, the evaluative process can genuinely transform the way decisions are made. For this to happen, impact evaluation must be closely linked to operations, involve the implementers from the very early stages of conceptualization, and evaluate programs delivered through the main institutional mechanisms available in a country. Such results are valid to the country context they are meant to inform and likely to be used by policy makers.

This way of thinking tries to meet the challenge of making research relevant to policy decision-making and the Bank's operational work.

² A classic example is from Neyman (1952) who computed a high and statistically significant correlation between stork nesting and baby births. Whether or not a correlation appears reasonable, however, has no bearing on the validity of the policy conclusion.

DIME History

The Development Impact Evaluation Initiative (DIME) was created in 2005 by the Bank's Chief Economist Office with the objective of generating knowledge on selected policies. Half a million dollars per year from the Bank's research budget was set aside to help the Bank conduct impact evaluations of policy interventions in multiple settings and use the evidence to produce generalizable conclusions about their effectiveness. Also in 2005, impact evaluation programs were started in the Africa Region and in the Human Development Network (HDN). These served as DIME's implementation backbone.

Expansion was rapid. In 2004, the Bank only had two dozen active impact evaluations (28). By 2008, that number had grown seven-fold (Figure 1). Through the Africa Impact Evaluation Initiative (AIM), the Africa Region's impact evaluation portfolio grew 40 times over so that today it represents 43 percent of active IEs (Figure 2). Similarly, through several thematic programs, HDN established a large number of impact evaluations and represents today 46 percent of active IEs (Figure 3).

Toward the end of 2008, the office of the managing director responsible for knowledge products and networks decided to mainstream and strengthen the role of impact evaluation in the Bank as a corporate priority. DIME's mandate was expanded to include two goals: institutional development for evidence-based policy and improvement in the quality of Bank operations. To ensure that the IE portfolio would reflect institutional learning priorities, DIME was redefined as a decentralized Bank-wide effort with a strong governance structure (see Annex A) and a coordinated approach to operational support.

Today, DIME has a high-level Steering Group composed of the chief economists and selected directors from all networks and regions. During 2009, the Steering Group met several times to set DIME's objectives and priorities, select priority thematic areas, and agree on an institutional structure conducive to enhancing ownership and quality. Within that framework, the DIME Secretariat provides technical leadership, networks maintain thematic responsibility, the regions are responsible for implementation, and the research and data groups in Development Economics (DEC) provide analytical and data services (Figure 4).

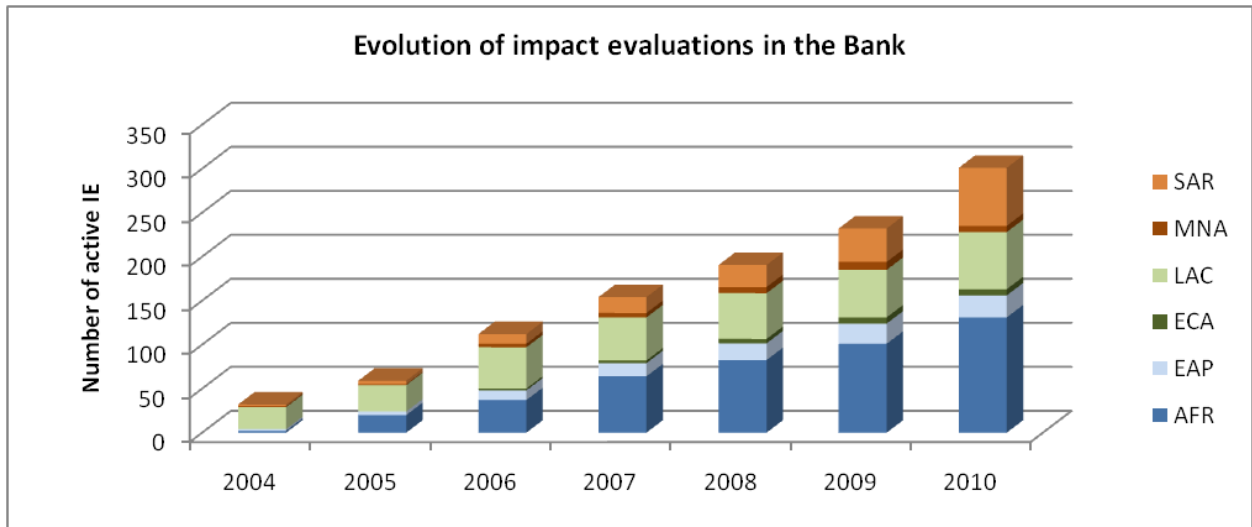


Figure 1. Evolution of Bank IE, 2004 - 2010

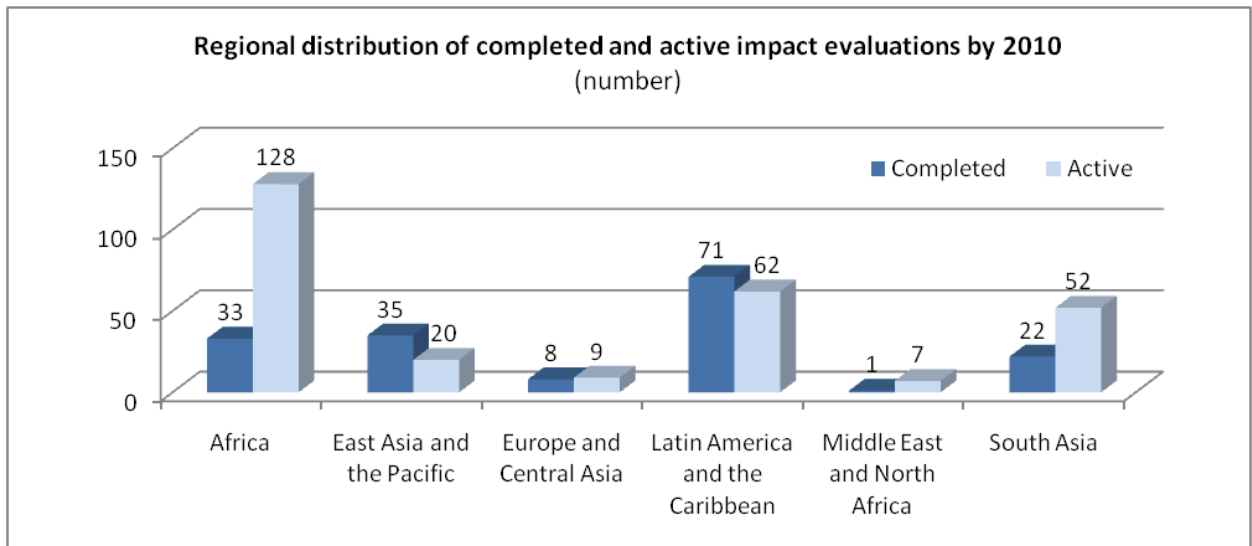


Figure 2. IE distribution by region

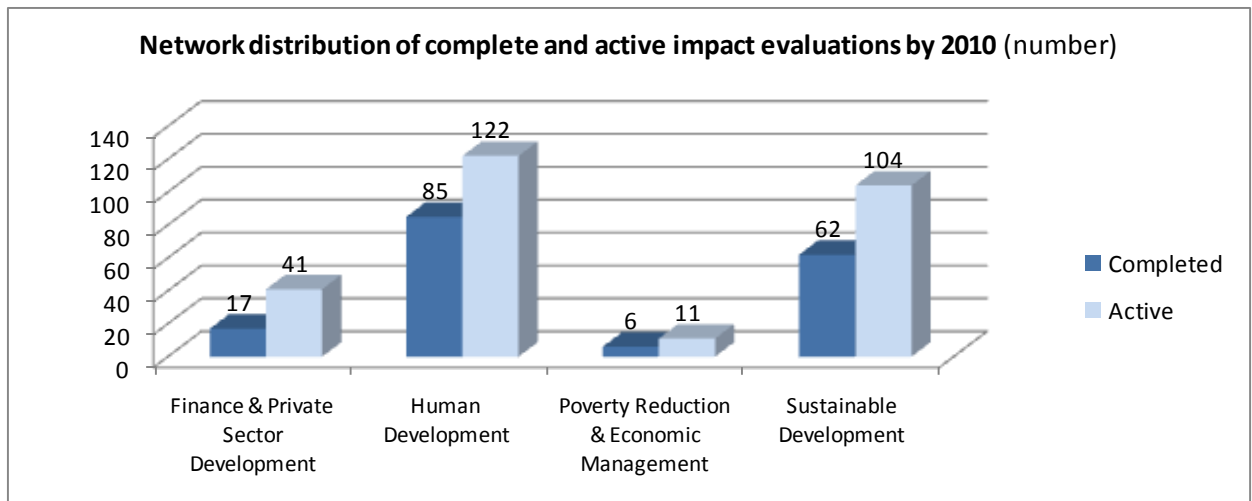


Figure 3. IE distribution by network

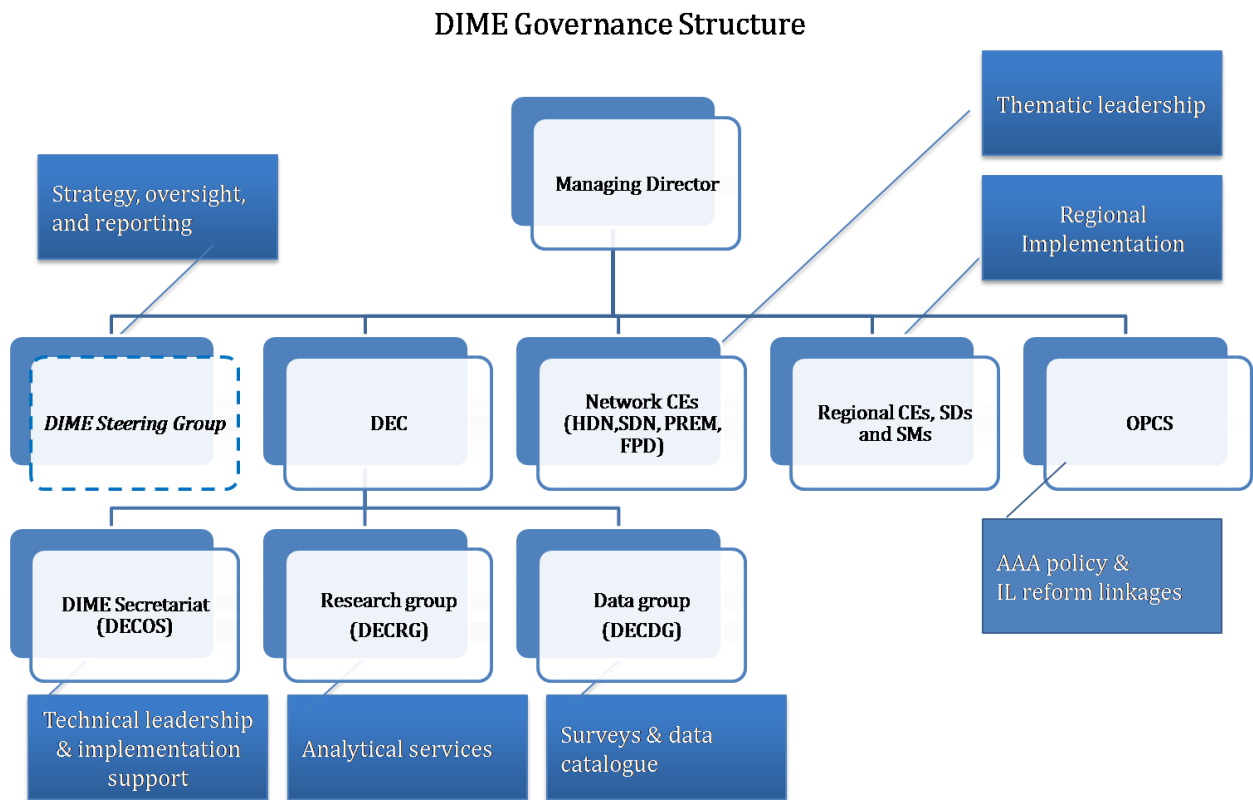


Figure 4. DIME governance structure

DIME Objectives

DIME has three objectives, as follows.

1. **Improve the quality of operations through iterative learning.** DIME works to integrate the Bank’s operational and analytical functions to incorporate evidence, empirically test the effectiveness of policy alternatives, scale up best-performing policies, and improve the effectiveness of programs during implementation.
2. **Strengthen country institutions for evidence-based policy making.** DIME builds in-country capacity to understand and use impact evaluation to inform policy and operational decisions. Modalities include training, networking, and learning-by-doing via joint Bank-government evaluations.
3. **Generate knowledge on critical development questions.** DIME seeks to secure the validity of learning for the country in which it operates by working with government programs at scale and in the prevailing institutional environment. By working programmatically across many countries and institutional settings, DIME also works to extract broader lessons of global interest.

DIME Strategy

Step 1: A results-based product to improve operational quality

“Basic characteristics of an individual organism: to divide, to unite, to merge into the universal, to abide in the particular, to transform itself, to define itself, and as living things tend to appear under a thousand conditions, to arise and vanish, to solidify and melt, to freeze and flow, to expand and contract. Since these effects occur together, any or all may occur at the same moment.” -- Johann Wolfgang von Goethe (1749 - 1832)

DIME promotes a results-based model for the Bank’s operations. The model strives to reinforce the analytical content of operations from design to completion by having research work with operations *throughout* the project cycle (World Bank 2006). Traditionally, operations and research functions are separate and communication between functions is not always easy. When researchers report their results, sometimes operations staff are not aware (Box 1). For this reason, DIME works to transfer how-to technology that goes beyond the teaching of analytical methods. The idea is to facilitate a whole process of collaboration among policy makers, the Bank’s operations, and internal and external researchers.

The model is aligned with the Bank’s results agenda and contributes to strengthening results-based products. The Bank traditionally conceives of a project as a chain of inputs, activities, and outputs designed to achieve a set of outcomes. This *results chain* is summarized in a results framework that defines the project development hypothesis. All projects are required to monitor their progress by using trend analysis, or before-and-after *impact assessment*, and report their findings in their implementation status reports, results reports, and implementation completion reports. In addition, the Results Platform defines a set of common indicators for the purpose of Bank-wide reporting. The monitoring approach is required to provide a basic accounting of Bank operations. As long as it is used to report project outputs (roads, schools, bridges), the approach is necessary and methodologically uncontroversial.

Box 1. Education in Punjab, Pakistan

A Bank research team spent 5 years analyzing the impact of the Punjab education sector reform supported by a Bank project. By 2008, the research had uncovered an exceptionally rich set of lessons. The bulk of the increased enrolment in Punjab came in the private sector (which the reform did not support) and the increase in public sector enrolments was identical in other provinces which were not subject to the reform. There was little evidence that school management councils helped improve outcomes, and, at \$400 per additional pupil enrolled, girls’ stipends were very costly (Andrabi et al. 2007). The research suggested that the project focus on the governance agenda, especially on public-private partnerships.

When the concept note for a phase 2 operation was distributed, the research team suggested several changes to the proposed activities on the basis of the analytical results. While the research team and the project team have diverging assessments on how much of the results was incorporated in the phase 2 operation, a review of the exchange suggests that more can be done to ensure that relevant knowledge is incorporated into Bank support. Creating the environment for fruitful exchange is one of the aims of DIME.

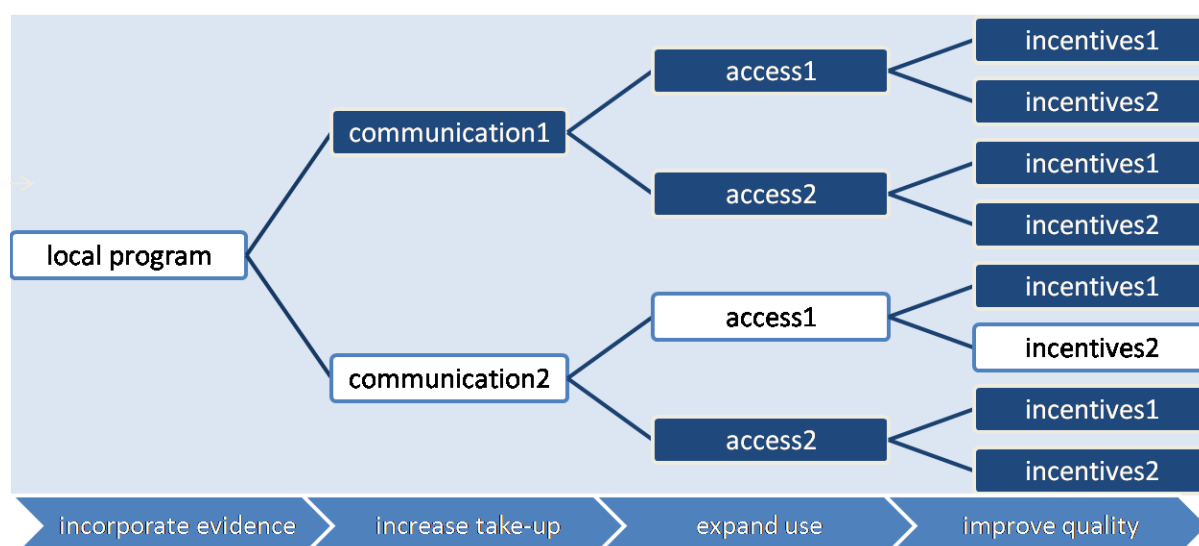


Figure 5. Test nodes of decision tree to improve program effectiveness

Methodological problems arise when monitoring is used to infer the contribution of a project to changes in developmental outcomes. This is because outcome trends are affected by many factors, and trend analysis cannot determine either the size or direction of a project's contribution to outcome changes. To infer attribution or causality, more analytical structure is required. By imposing analytical structure, impact evaluation can be used to measure project effectiveness and validate its development hypothesis. This is referred to as *summative* evaluation because it provides an overall view of whether the project worked to deliver the desired results.

In addition to bringing impact evaluation into the results agenda, DIME aims at introducing *formative* evaluation based on experimental methods. Formative evaluation compares alternative mechanisms within a project to discover how best to implement the project. By doing formative evaluation, DIME aims at affecting the quality of operations in real time. This requires a change in the way projects are conceived to allow for competing hypotheses to survive side-by-side during project design and implementation until the empirical evidence proves their relative effectiveness. *Decision trees* are used to describe the competing hypotheses the project is considering.

Formative impact evaluation tests the nodes of these decision trees to compare alternative communication strategies to secure higher take-up, to compare access strategies to secure greater use, and to compare supply incentives to improve quality. By applying experimental methods to formative evaluation, DIME helps projects produce scientifically valid guidance for steering projects toward better operational alternatives and greater overall success (Figure 5). DIME's idea is to cater to the demand for *impending* knowledge, knowledge needed by policy makers today to make upcoming decisions. This knowledge must be actionable, generated by and valid for the context for which it is meant and geared toward improving results.

During the December 2008 workshop in Dakar, for instance, the Senegal HIV/AIDS agency laid out its plan for rolling out its new HIV prevention strategy (peer counseling) to substitute for the old strategy (social mobilization). During the clinic, the government

modified its plans: the new strategy would be randomly phased-in in some health districts to measure the performance of the new strategy relative to the old. In record time, by January 2009, the government randomly assigned one-third of the health districts to one of the following: (i) traditional social mobilization; (ii) peer mentoring; and (iii) no intervention (Sakho 2009). A year later, analysis of administrative data confirmed why a program would want to introduce changes in this fashion. By doubling the number of people who get tested and pick up their results, peer mentoring proved to be more effective for HIV-positive individuals, while social mobilization proved to be more effective in increasing the number of partners of HIV-positive individuals who come and get tested (Arcand, Sakho, and Wagner 2010). The two interventions are not substitutes but rather complement each other in reaching out to different populations at risk. The findings are of interest to both Senegal's and global HIV prevention efforts. Another example is the Descomplicar project in Minas Gerais, Brazil, where the evaluative process is thought of as a way to secure the success of a large reform program that simplifies business registration and establishes one-stop-shops across the state (Box 2).

DIME's approach to supporting the results agenda and improving the quality of operations includes the following elements: (i) sharing the existing evidence at the design stage to introduce innovative and effective solutions to development; (ii) experimentally testing project modalities to improve results; and (iii) validating the project's overall development hypothesis. While (ii) is not always present in DIME's evaluations, it has risen from 11 percent of completed to 39 percent of active impact evaluations.

The challenges associated with this process include: (i) the lack of flexibility of traditional lending products (Does the project need to be restructured when it changes course or can these changes be included in project design?); (ii) the difficulties in effecting cultural change (Should we admit to the government we do not know what works best?); (iii) the adverse incentives to evaluating one's project (Will my project be rated unsatisfactory if the evaluation shows no results or satisfactory because we help avoid similar errors in the future?), and (iv) the time it takes to secure the agreements and gather the financing.

These elements and challenges go to the heart of the results-based product the Bank is trying to define: a product that flexibly adapts to new information and evolves to secure greater effectiveness. The sharing and learning is delivered through intensive Bank technical assistance to secure continuity and longer-term commitment. Much of DIME's effort goes into working with staff, management, government officials, and researchers to help internalize the value of impact evaluation. To protect against adverse incentives, DIME stresses the evaluation of competing alternatives to improve results, and raises funds to subsidize the evaluations. DIME invests in helping individual researchers work with government partners by hiring coordinators to lower transaction costs, building capacity to lower communication barriers, and raising funds to finance advisory services and field presence.

Aligning the Bank's incentives with a results-based culture will require making its lending products more flexible, including learning as a project objective when relevant, and strengthening managerial incentives to reward learning.

Box 2. Fine-tuning Brazil's Descomplicar

The Government of Minas Gerais in Brazil participated in the World Bank Workshop for Finance and Private Sector Impact Evaluation held February 1-4, 2010, in Dakar, Senegal. Representatives included staff from Descomplicar, a project to simplify business procedures across the State of Minas, and colleagues from Junta Comercial, the local chamber of commerce.

This opportunity was of great value to us. We learned about the methods for evaluating government programs. We were offered access to Bank and international experts who worked with us to develop an evaluation for one of our programs. We interacted with colleagues from many different countries and engaged in similar reform processes, such as the team from South Africa's Department of Trade and Industry with which we exchanged notes on the one-stop-shop reform process.

We used the opportunity to design the evaluation of the Minas Program to Reduce Informality. What we discovered in this process was that we needed to think even harder than we had before about the mechanisms to ensure that our reform would succeed. We have already implemented a one-stop-shop reform in 29 municipalities. How could we take full advantage of this reform for the purpose of reducing informality among small firms?

The first question we decided to tackle was whether in fact registration has a net benefit for small firms. Are the benefits of registration greater than the costs associated with registration fees, mandatory accounting practices and taxation? The second question we set out to respond to was whether it was lack of interest or lack of information that keeps small firms away from formality, and what kind of communication strategy and incentives would be most effective in getting small firms to register their business. Third, we wanted to find out how effective was our policy of punitive enforcement as implemented by a corps of inspectors.

Our reform has not changed its approach and objectives, but what the process of planning the impact evaluation did was to help us fine-tune our interventions to make them more effective.

The work did not stop in Dakar. Sustaining this effort is viewed as something important not only by our team but also by our Secretary of Planning. With her mandate, we established a working group that today includes five separate institutions, the World Bank team, and local researchers to move to implementation. We expect that the results from the evaluation will help us steer the program in the right direction.

Priscila Malaguti and Joao Luis Soares, Projeto Descomplicar, Planning Secretariat of the State of Minas Gerais, Brazil

Step 2: Building capacity through joint evaluations

DIME's approach places policy makers and program managers at the center of the evaluative process. DIME brings countries the support needed for them to find their own solutions. The approach is mindful of country ownership and the uniqueness of local institutions.

DIME's sustained effort is the key to its capacity development strategy. Elements include: (i) formal training, (ii) networking with a large community of practitioners, and (iii) learning-by-doing through joint government-Bank evaluations. The medium-term objective is to help policy makers develop understanding of the tools put at their disposal and their ability to take ownership of a more evidence-based

Box 3. “We can make a real difference in future policy developments”: South Africa’s National Department of Human Settlements

In 2004, the National Department of Human Settlements of South Africa initiated the Informal Settlements Upgrading Programme (ISUP) under a new and broader Human Settlements policy called Breaking New Ground. The main objective of ISUP is to facilitate the structured upgrading of informal settlements and improve the living conditions of least advantaged populations. One of the functions of the National Department would be to monitor progress on housing policy, such as the number of housing units delivered and the level of access to essential services.

After participating in a World Bank workshop on impact evaluation held in Pretoria in 2006, the Department embarked in a long-term collaboration with the Africa Impact Evaluation Initiative at the World Bank to improve the M&E function of the department. As a result, NDHS’s Chief Directorate on Monitoring and Evaluation launched a program of rigorous impact evaluation using innovative ISUP interventions in the provinces of Gauteng, Limpopo, and Free State in 2008.

The program seeks to estimate the impact of ISUP on empowerment, health and safety, employment, consumption, and productive activities. Studies have been started in three provinces and everyone has been learning in the process. We contracted the survey company, helped define samples and questionnaires, and participated in the training of the trainers’ workshop on data collection in early 2010. We conducted the pilot study in Atteridgeville and the household listing survey in Free State and Limpopo provinces. Training of the field workers was completed in the Free State in January and in Limpopo in March. The Household survey was started in April and completed in May. Data analysis and report writing will follow.

The implementation of these studies is enabling NDHS to evaluate the effects of its interventions on the lives of targeted populations and find the common ground with other national departments. Collaborations started with the Departments of Health, Education, Water Affairs, and Transport. With the Department of Health we have a formal arrangement to evaluate impacts on anthropometric, anemic, and cognitive development indicators.

The whole process has meant a transformation in how we think of our role as a national department in supporting the improvement of operations in the provincial governments in housing and across sectors. The difference is that today we feel we can count with reliable evidence, whereas before we could only use suggestive evidence. This is encouraging us to think that we can make a real difference in future policy developments and enhance our programs and those of provincial governments on the basis of what the evidence says.

Due to this, the National Department of Human Settlements plans to roll out the impact evaluation program to ISUP interventions in Eastern Cape, Western Cape, Mpumalanga, and KwaZulu Natal provinces.

Mulalo Muthige, National Department of Human Settlements, Republic of South Africa

approach to development. The National Department of Human Settlements of South Africa, for example, used the experience to reposition the department in the context of their national housing policy and provincial housing programs (Box 3).

The effort is sustained through periodic workshops that provide clients and Bank staff with a forum to compare and benchmark their results and learn from the experience of others. Clients become members of a cross-national club of peers with access to international experts from the World Bank and many partner institutions. The model for these workshops was first tested in Mombasa in June 2005 and perfected over time. The workshops target impact evaluation training to project teams and combine it with clinics where each team designs its impact evaluation. The core delegation should include a director- level official from the relevant ministry (policy-making); a program manager

(knowledge of the intervention); and an economist/statistician (follow-up). The delegations are trained in impact evaluation methods and exposed to international IE results relevant to their sector. The clinics, facilitated by IE expert, stimulate discussion around the evidence and nurture a process of critical thinking aimed at defining a learning agenda. On the last day, each delegation presents its designs for plenary discussion. Delegations go home with a written product that they can present to their own ministries. Once delegations confirm their intention to do the impact evaluation, the country is provided the technical assistance needed for the implementation.

The same teams meet again on an annual basis to share evidence and experience. For example, in the Africa Program for Education Impact Evaluation (APEIE), country teams met in 2007 to plan evaluations, then at the end of 2008 and again in May 2010 to share knowledge and begin to extract lessons (Box 4).

Box 4. Africa Program for Education Impact Evaluation (APEIE)

Started as a partnership between the Africa Impact Evaluation Initiative and education team, APEIE was established in 2007 to improve the ability of countries to meet the MDG goals in the context of the Education for All Fast Track Initiative, which has been financing the program. APEIE objectives are to: (i) build technical and organizational capacity in the education sector to conceptualize and implement rigorous impact evaluations through a learning-by-doing approach; (ii) build country-level evidence on the effectiveness of education policy and programmatic alternatives; and (iii) provide cross-country venues for dialogue, networking support and publication outlets to foster cross-fertilization and peer-to-peer learning regarding the efficacy of education interventions.

APEIE has been supporting 14 country-specific impact evaluations and convening regular cross-country meetings to build capacity and discuss policy lessons.

- In Abuja in 2007, the country delegates and the team task leaders received training on impact evaluation and international evidence on education policies. Each team applied their new knowledge to the development of the impact evaluation for their education program. Ghana was interested in evaluating school management committees and Senegal was interested in school grants.
- In Dakar in 2008, countries presented their baseline results and discussed challenges moving forward. Already, the workshop was in the hands of the participants sharing notes and experiences. It was a good opportunity to hear from clients and team task leaders alike.
- In Accra in May 2010, the original countries and new recruits into the program discussed how the evidence guided their education programs and what new policy innovations they are considering moving forward.

Today APEIE is the first program to be fully integrated into a sector unit and mainstreamed into the operations. Two economists are based in the education unit and provide full-time support to operations to incorporate a results-based approach in the Africa education sector portfolio.

Arianna Legovini, Muna Meky, and Jee-Peng-Tan (2007)

Box 5. “We won’t be passive passengers”: Malawi’s evaluation of peer farmers

“We don’t want to be passive passengers in the study, and just receive the results at the end. We want to be involved and learn together.” -- Malawi’s Deputy Director of Agricultural Extension, Ministry of Agriculture

The Ministry of Agricultural and Food Security has taken full ownership of the impact evaluation of the Malawi Agricultural Development Program. None of the team involved with the impact evaluation has experience with randomization or rigorous evaluation methods, but they are eager to learn. Rather than rely on an external consultant, the Ministry is keeping as much of the impact evaluation work as possible in-house, in order to understand and learn from each step of the process. At the request of the M&E Unit, I have created a variety of capacity building exercises, to ensure that the team can fully participate in the impact evaluation, and that they will have the skills needed to carry out similarly rigorous evaluations in the future.

Below is a summary of some of the capacity building activities completed in the first year of the impact evaluation:

Introduction to CSPro: A basic introduction to the program and its uses for various MoAFS staff.

Using CSPro to create data entry templates: A hands-on workshop for staff from the Department of Agricultural Planning Services. After the workshop, each participant completed a ‘practicum’ by designing one page of the data entry template for the baseline household questionnaire.

Using CSPro for data entry: A team from a local agricultural college was hired to do the baseline data entry, with training and supervision provided by me and the MoAFS team. We spent four days training Bunda College staff and data entry clerks in the use of CSPro and familiarizing them with the baseline data template.

Exporting data from CSPro to Stata: Once data entry was complete, I conducted a joint training for MoAFS and Bunda staff on how to export the data from CSPro and how to write a program to merge and reconcile the double-entered data in Stata.

Data management and analysis with Stata software: A three-day workshop, organized in collaboration with IFPRI. The training combined econometric theory with practical instruction in how to use Stata; we covered nine topics, ranging from importing and exploring data in Stata to analyzing survey data. Participants included staff from the M&E unit and the Department of Trade and Marketing at the Ministry of Agriculture, IFPRI, and the Millennium Challenge Account (local affiliate of the MCC, which is supporting the ADP-SP impact evaluation). Although most participants were familiar with data analysis, few had ever used Stata before, so the training aimed to create a foundation level of understanding for the core team that will be involved in management and analysis of the ADP-SP impact evaluation data.

Follow-up Stata workshops using ADP-SP baseline data: In the next two months, we are planning a series of workshops, each of which will be a hands-on application of the topics covered in the first Stata training, using the ADP-SP baseline data set. The principal investigators for the impact evaluation will participate in the final sessions on data analysis, and the end result will be a collaboratively written baseline report, to be presented to Ministry of Agriculture staff at a national and district level.

Maria Jones

ADP-SP Impact Evaluation Field Coordinator, Malawi

In-country capacity development is based on learning-by-doing through joint evaluations. Activities include: (i) in-country workshops and training on data analysis and (ii) a Bank-financed research team and full-time field-based coordinator in charge of analytical quality, field supervision, and support to implementation, M&E, and data collection. Most activities are conducted jointly with the implementing agency's team: aligning IE and the M&E data collection strategy, developing data collection instruments, preparing the terms of reference for data collection, training the enumerators, and supervising data collection. The joint work builds capacity in the implementing agency and in the data collection agency. The new DIME Microdata Catalogue will further enhance this process by clarifying how proper data documentation is kept to ensure data usability and availability (Annex D).

When the data is collected, the research team trains the local counterparts in data analysis using software that includes all the main routines required for impact evaluation and other economic analysis. This is to make sure that the local team can make full use of the data for the purpose of the evaluation, the monitoring function and beyond. The example of Malawi shows the type of training that is delivered in practice (Box 5).

When the results are estimated, the implementing agency is the first to know. This is important for internalizing the value of IE and contrasts with a more common model of research where the implementing agency is usually unaware of the results even after they are published. For example, the Ministry of Education of Madagascar was involved in the analysis and took the decision to scale up the school management program when the preliminary results were out, more than a year before the evaluation report (Box 6). Direct involvement helps develop a broader understanding at different levels of government regarding the use and applicability of evidence for guiding decisions. By supplementing the evaluation with costs estimates, government can obtain cost-effectiveness measures. These are of great interest to the planning and budgeting functions of government, which are often severely constrained in their ability to allocate financing efficiently.

Challenges to implementing this model of capacity development include: (i) maintaining the medium- to long-term commitment required for institutional development while working with shorter-term instruments and financing; (ii) reaching a critical mass within a country; and (iii) the generally low capacity in many local academia.

To strengthen capacity of local academia, DIME's effort is being complemented by others. IEG's CLEAR initiative aims at developing Anglophone centers of excellence on evaluation, including impact evaluation. CEQA, based at the University of California-Berkeley, focuses its effort on local research institutes to build capacity for impact evaluation. The Agence Française du Développement is currently developing a program to build impact evaluation capacity in Francophone research institutes. These important initiatives aim at improving the local markets for evaluation complementing DIME's focus on building government capacity. Together they can improve the demand and supply of impact evaluation services.

Box 6. Building capacity for Madagascar's improvement of school management

The Malagasy government fully participated in the evaluative process of Madagascar's Improvement of School Management program (AGEMAD). The evaluation originated in an education sector analysis that revealed the management weaknesses in the education system. The Ministry of Education assembled a team of educators to examine the options for renewing and rationalizing the tools for school management. The team worked with sustained support from the Bank team to elaborate a new set of tools for educational management focused mainly at the school and classroom levels. They consulted extensively with their colleagues in regional offices and with school heads and teachers, a process that yielded a new set of tools that was both validated and owned by practitioners in the system.

The ministry officials were keen to test the efficacy of the new tools before scaling them up country-wide. Thus, the idea of a rigorous impact evaluation took root. In preparation for the impact evaluation, the Bank team facilitated the Malagasy team's participation in multiple training workshops, in-country and elsewhere, and arranged for their active engagement with external technical experts to conceptualize the design of the evaluation and the various questionnaires for data collection.

The Malagasy team took a leading role in implementing the interventions to be evaluated. The dedication of the 15-member Malagasy team made it possible for the team to adhere closely to the exacting timing and process protocols required in the experimental design, with regard to both implementation and data collection.

The partnership built capacity. One indication was that the Malagasy leaders understood the substance of the impact evaluation intimately, even if they needed help on the technical front. They were confident in presenting this work to their peers and members of the local education donor group. Another indication was that when the succession of two technical assistants hired to help supervise the implementation had to leave after 18 months, the Malagasy leaders judged that the team had learnt enough to be able to carry on without outside assistance. The work was completed smoothly for the remaining months of implementation.

When the results of the impact evaluation were shared and discussed with the Malagasy team, the relevant government officials made the decision to disseminate the tools more widely, a move made more than a year before the evaluation results were published in a report (World Bank 2010a). They integrated the AGEMAD management tools into the curriculum of teacher training programs and encouraged schools to use the tools to train the large number of community teachers funded by Madagascar's grant from the Education for All Fast Track Initiative Catalytic Fund. More generally, the AGEMAD impact evaluation gave Malagasy policy makers an appreciation of the benefits of impact evaluation. The same team launched another impact evaluation in 2008, this one focused on school feeding.

Jee-Peng Tan, Advisor, Education Department, Human Development Network

World Bank, *Améliorer la gestion de l'enseignement primaire à Madagascar - Résultats d'une expérimentation randomisée*, Africa Human Development Working Paper Series (Washington, DC: forthcoming, 2010).

Step 3: From local learning to global learning

With 170 completed and 280 active studies in 72 countries, DIME is the largest program in the world designed to systematically learn from development experience on the basis of rigorous impact evaluation (Figure 6).³ It is also one that leverages the largest network of established relationships with governments across the developing world. DIME's cross-country scope helps extract global lessons learned. The process of generating knowledge will strengthen the position of the Bank in the economic development debate and its relationships with external clients over time.

DIME focuses on priority policy areas that are expected to drive the development agenda in the next decades. The dearth of concrete evidence and large knowledge gaps motivate DIME's work. Together with its many academic partners (102), DIME targets knowledge gaps to make a contribution to both Bank research and global knowledge. The work is helping the Bank and its clients understand what policies work, for whom, under what conditions, and in what type of institutional environments. They are helping to shape the environment for greater economic growth and poverty reduction.

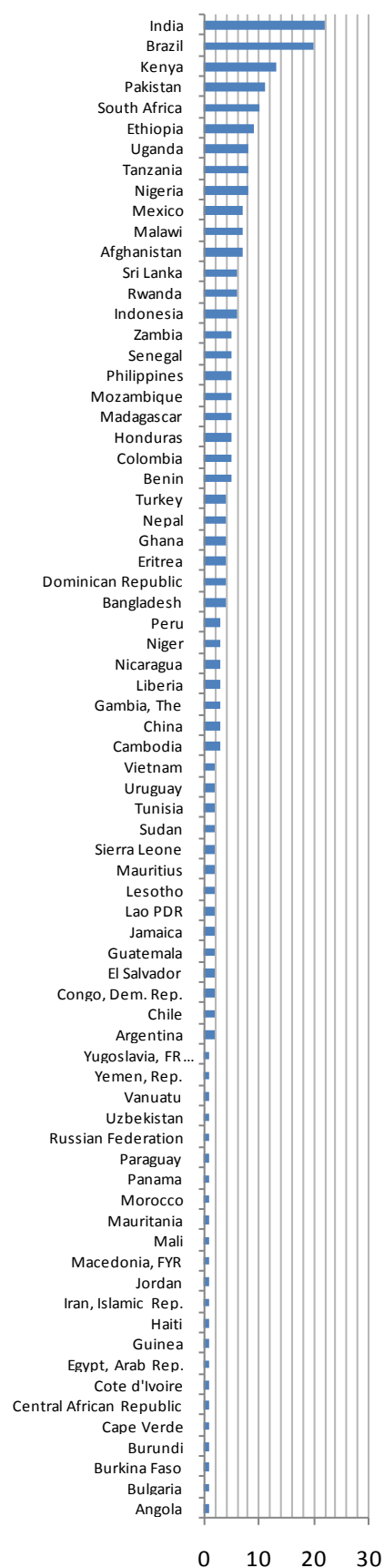
The selection of policy areas reflects a Bank-wide consultative process conducted from December 2008 to March 2009. The DIME Steering Group decided to expand the DIME program which was active in local development, education, health, social protection and infrastructure and include early childhood, private sector and finance, adaptations in agriculture, water and forestry, mitigations in energy, and institutional reform, in addition to gender and fragile states as cross-cutting themes. Combined, these programs are generating a large body of lessons and guiding principles for development work.

The questions addressed are of strategic and corporate interest and have strong potential for cross-country applicability. They include, for example, the following:

- What are the structures of accountability that improve the quality of education?
- What are viable low-cost early childhood interventions that can help level the cognitive playing field?
- What performance-based incentives substantially improve health outcomes?

³ This compares for example with 86 completed and 86 ongoing studies by MIT's Poverty Action Lab. Thirty six of DIME's active IEs are joint with JPAL's associated researchers.

Figure 6. Active Impact Evaluations



- What type of HIV and malaria prevention helps halt the epidemics?
- How to rapidly increase the rate of technology adoption to secure food security and high returns to rural infrastructure?
- What incentives for energy efficiency and environmental protection enable climate change adaptation and mitigation?
- What regulatory reforms release private sector potential and increase growth?
- What accountability practices strengthen governance and local business development?
- What institutions create a more accountable and effective local government?
- What combination of basic infrastructure contributes to breaking the cycle of poverty?

DIME works programmatically and under a common framework of analysis and measurement. Each program develops a research agenda and the data instruments and indicators that will be used. For example, a brainstorming with a large gathering of political scientists and economists (Experiments in Governance and Politics or EGAP) was used to understand what critical policy questions should frame the fragile states program. DEC's Agricultural Living Standards Measurement Surveys (LSMS), developed for large country-representative panel data, will be used for the AADAPT program to ensure data comparability within and across countries and the ability to benchmark results. *Making the Grade*, a comparative study of education achievement across African countries (Beltran et al. 2010) was made possible by APEIE, which conducted the same cognitive and achievement test in several African countries.

The approach helps synthesize results. The syntheses extract generalizable conclusions from country-specific knowledge, and model policy response. The report *Conditional Cash Transfers: Reducing Present and Future Poverty* (Fiszbein and Schady, 2009) completed last year provides a good example of the synthesis reports envisioned by DIME as the primary way to summarize the broader lessons from programs of impact evaluations. Drawing on impact evaluations from a dozen countries, the report not only documents the evidence of impact of CCT programs but also provides important lessons on a range of design issues.

Forthcoming are the reports on education accountability (World Bank 2010b) and local governance (Mansuri and Rao 2010). The education report will present the evidence on the impact of school accountability interventions in improving learning outcomes. School-based management, the decentralization of decision-making to school-level agents, the use of actionable information by school administrators, teachers, parents and students, and the reforms of teacher contracting and pay are some of the interventions that will be covered (Bruns, Filmer, and Patrino 2007). The local governance report will summarize the impact of interventions to decentralize policy making to local governments and communities, enhance participation and inclusion of traditionally excluded populations, and empower them through information, organization, and resources. In addition, IEG has been using IE results to compile a summary on nutritional interventions and one forthcoming on safety nets. These products enable the Bank to learn systematically and disseminate what works best in the design and implementation of development activities—a direct contribution to the Bank's knowledge agenda.

In addition DIME disseminates the results via cross-country workshops, seminars, conferences, network weeks, and web and journal publications. All IE results are published in the IE Working Paper Series and in the external IE database in a way that is accessible to all audiences. The new IE page of the World Bank website is currently under construction.⁴ During the DIME workshops, the results of impact evaluations (done by the Bank or others) are presented to the teams preparing new operations. The presentations are followed by project specific round tables that are facilitated by IE experts and that stimulate discussion around the evidence that is relevant to the design choices of those operations. The design of the average operation participating in DIME changes during the workshop. While figures supporting this claim are not currently available, DIME plans to track this indicator in the future and included as one of the issues for the DIME external evaluation. Furthermore presentations of IE results are made every year during Network Weeks to target network staff, and regularly in the AIM, DIME and HDN seminar series, to target the Bank and external audiences. IE results are also used to inform proposed sector strategies through direct inputs into the strategies or through DEC comments to the OVP discussions.

DIME Thematic Programs

“Theoretical ideas should always find important applications within the pupil’s curriculum. This is not an easy doctrine to apply, but a very hard one. It contains within itself the problem of keeping knowledge alive, of preventing it from becoming inert, which is the central problem of all education.” -- Alfred North Whitehead

The programmatic approach is DIME’s platform for sector learning. Multi-country thematic programs secure a coordinated policy-learning agenda and vibrant networks for generating ideas and knowledge. They provide policy makers access to top academic resources—a powerful interface to build capacity among our clients. They enable benchmarking of results and economies of scale in the analytical work and capacity development activities.

DIME sponsors 15 thematic impact evaluation programs (Figure 7). Of these:

- (i) Eight have been ongoing (education, health, HIV, malaria, CCTs, active labor markets, local development and access to infrastructure);
- (ii) Four are new (early childhood development, finance and private sector, institutional reform and agriculture) together with two cross-cutting themes (gender and fragile states; and
- (iii) Three are yet to be launched (forestry, water resource management, and energy mitigation).

The programs are described in Annex B.

Several of these programs have developed a coordinated approach across a multi-country context and assumed responsibility for programmatic and country-specific activities. The programmatic activities serve to adopt a coordinated learning agenda, ensure quality and build capacity. The country-specific activities serve to deliver specific analytical products.

⁴ The Impact Evaluation page of the World Bank will be available in the near future. It will replace IE pages from various units and departments.

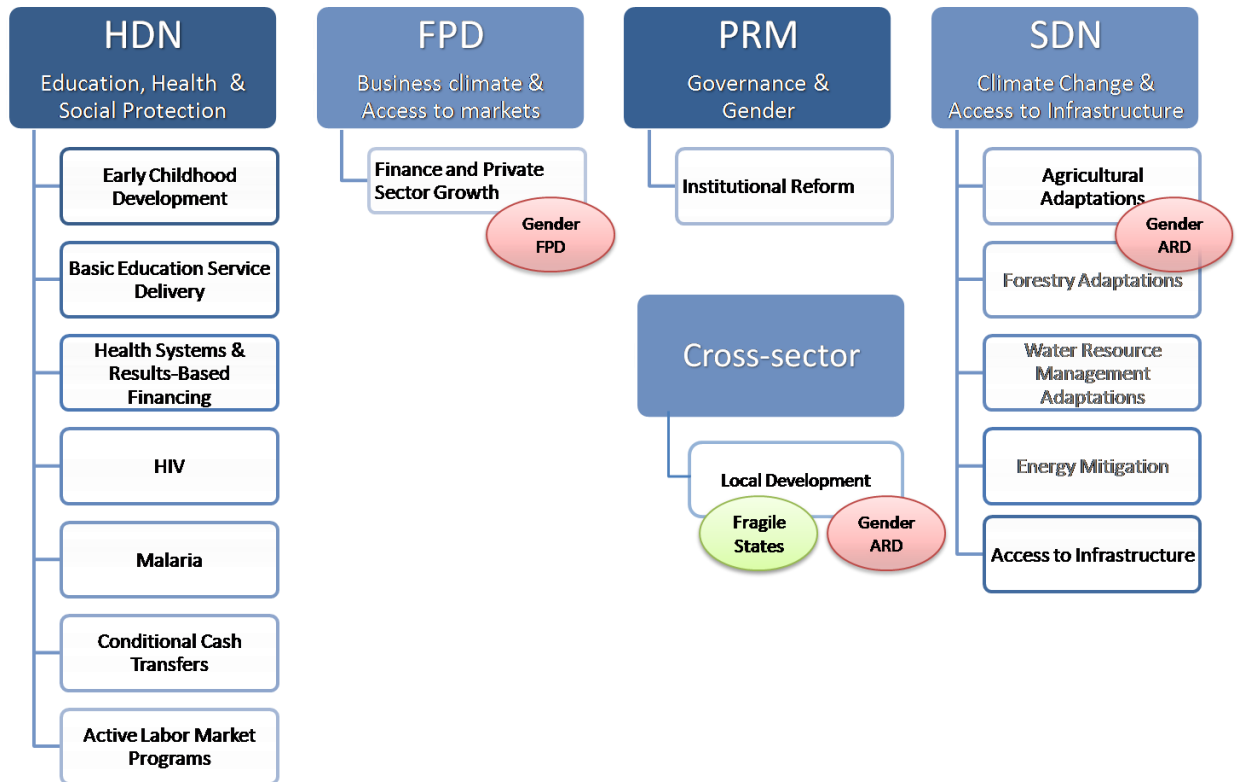


Figure 7. DIME thematic programs

Programmatic activities are targeted to all program participants. They include: (i) annual workshops; (ii) practitioner networks with access to experts; and (iii) a technical advisory group in charge of the analytical and measurement frameworks, delivery of training, quality assurance, and production and dissemination of knowledge products. Coordination is provided by one or two full time program coordinators—economists with strong impact evaluation skills—a sector lead, and a senior impact evaluation expert. Each program raises donor funds to carry out programmatic activities and finance advisory services in each participating country. Figure 8 illustrates these relationships.

In-country activities are targeted to each client. Technical assistance is delivered by a team of researchers and a field coordinator. The work is not contracted out but delivered by the Bank with the support of internal and external researchers. As a result the Bank has now unique expertise in this area. The research team works with the project team and government counterpart throughout the design and implementation of the analytical products, including preparation of data instruments, supervision of data collection, monitoring of intervention and data analysis.

Program development and project selection. DIME tries to implement programs as a partnership with sector management. Good examples are the management of the Africa Region’s education, health, agriculture, and private sectors, the Latin America agriculture and private sectors, Brazil CMU, the Poverty Reduction and Economic Policy Network (PREM) in the Middle East and North Africa Region (MENA), and South Asia’s rural

livelihood team. Partnerships with sector managers and directors are critical for changing the culture of evidence-based policy in the Bank and among clients.

Applicability. The intervention that is evaluated does not depend on the financing mechanism. DIME evaluates Bank-supported interventions or reforms that are financed through an investment loan or supported within the framework of a sector-wide approach or development policy loan. Sometimes, at the request of the client, DIME will evaluate interventions that are not Bank-financed.

IE coverage. The scale and level of project coverage is a management decision coordinated with task team leaders and their counterparts. Coverage varies widely. Overall, 13 percent of active projects have an impact evaluation. Regionally, coverage ranges from 19 percent in Africa and South Asia to 2 percent in Europe and Central Asia (Figure 9). By network, 26 percent of HD projects have impact evaluation and 4 percent of PREM projects (Figure 10). Newer projects have greater coverage. DIME-FPD covers 75 percent of Finance and Private Sector (FPD) fiscal 2010 projects in the Africa Region and some FPD projects in other regions. AADAPT covers 20 percent of the Africa Region’s fiscal 2010 projects, 25 percent of the South Asia Region’s, and 33 percent of the Latin America and Caribbean Region’s.

Because projects self-select into impact evaluation, the evaluated projects may not be representative of the Bank’s portfolio. This may affect the perceived effectiveness of the Bank’s projects if the better projects volunteer for evaluation. This problem is diminishing in importance as impact evaluation moves to capturing a larger proportion of the relevant portfolio.

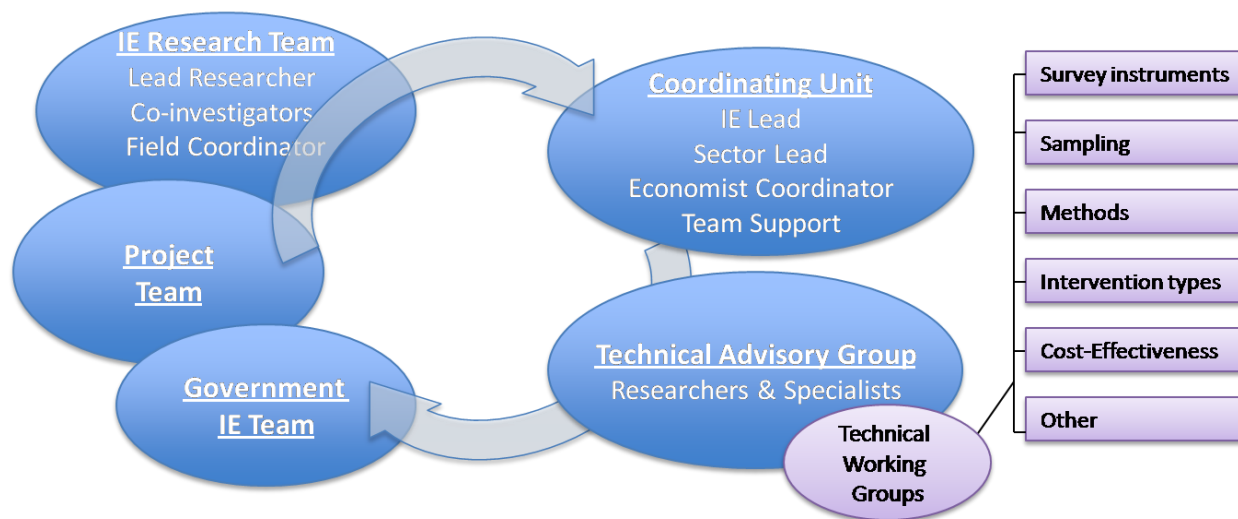


Figure 8. Organization of a thematic impact evaluation program

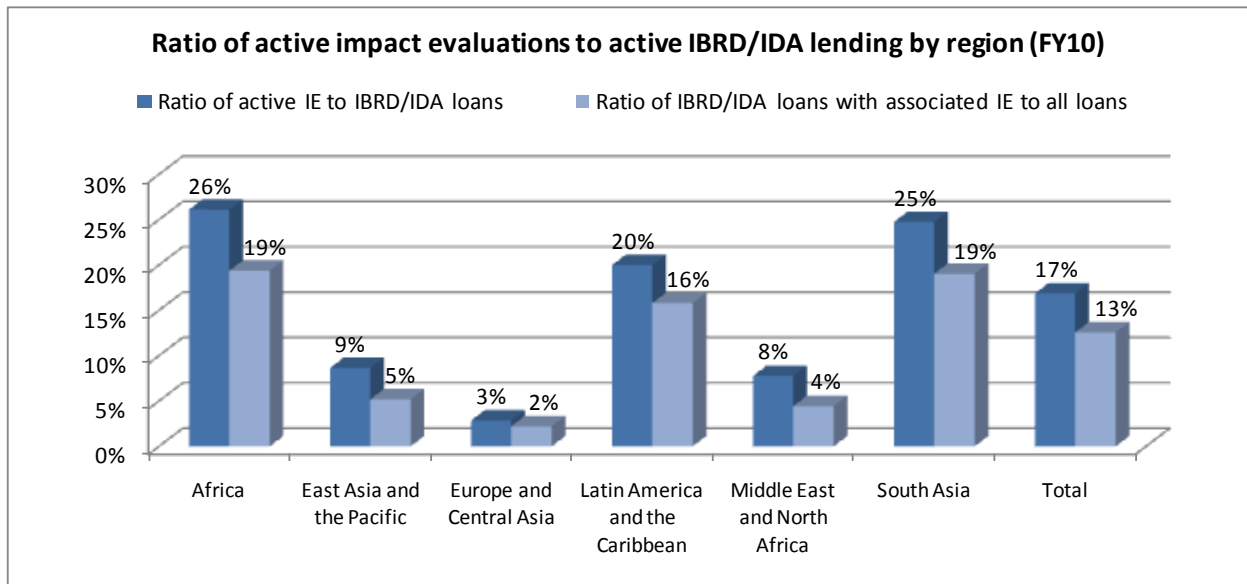


Figure 9. IE in lending by region, FY 2010

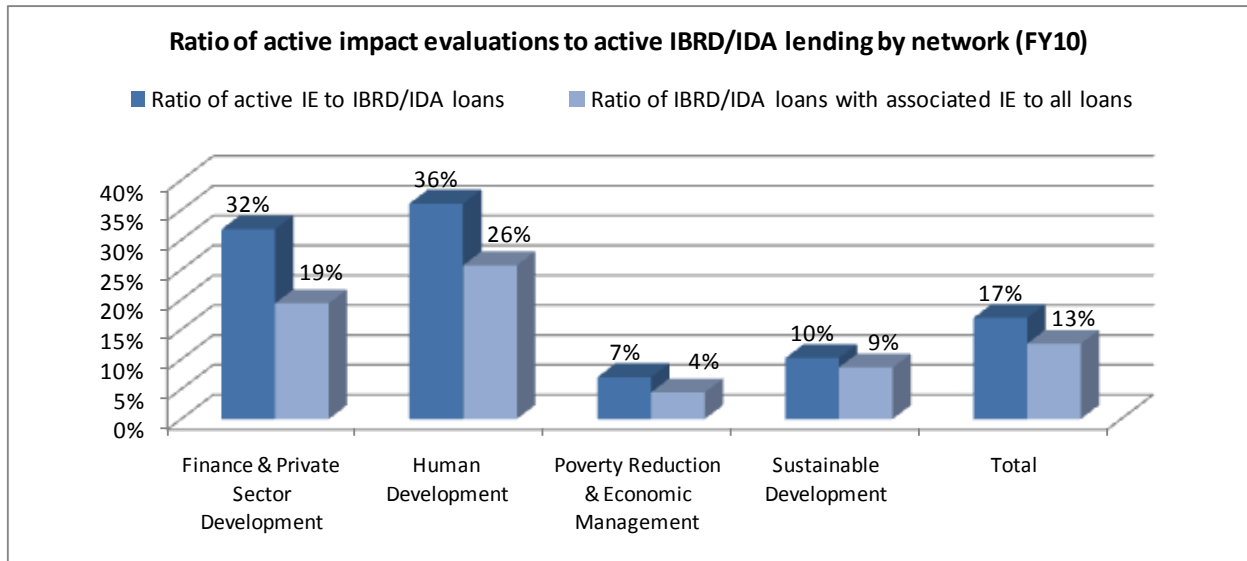


Figure 10. IE in lending by network, FY 2010

Quality assurance. The IE code was created in 2006 to recognize impact evaluation as a separate AAA product. The IE code requires approval from country director and sector manager, concept note, peer review, key milestones, publication and dissemination, and activity completion summary (World Bank 2006). While currently ongoing evaluations have not consistently followed this practice, it is DIME’s policy to increase the proportion of impact evaluation with an IE code from current levels (42 percent⁵, Figure 11) and ensure that the process is followed. This is not without difficulties but necessary to make IE products accountable to the country and sector teams.

Many evaluations also undergo further scrutiny and review as part of their efforts to raise funds. Within the Bank, the Spanish Impact Evaluation Fund conducts a full technical

⁵ This figure includes other analytical product codes linked to an IE.

review of the proposed activities, ranks proposals, and approves funds allocation through a technical committee discussion (www.worldbank.org/sief). The approval process for the Education Program Development Fund (EPDF) accompanies and supplements the Bank’s review process. Outside the Bank, the International Initiative for Impact Evaluation (3IE) initiative follows a competitive process for awarding research grants.

Thematic programs offer additional quality assurance provisions. The Technical Advisory Group of each program functions as a sounding board developing a common analytical framework, agreeing on quality instruments for data collection and conducting prior internal review of concept notes.⁶ Data quality is secured through the development of common data collection standards, terms of reference and instruments. DIME has been supported by DEC’s LSMS team which has unique expertise in the area of household surveys (<http://www.worldbank.org/lms>). Data access is supported by DEC’s Data Group which helped DIME establish a data catalogue. Impact evaluation data will be made available with supporting documentation for usefulness and usability (see Annex D).

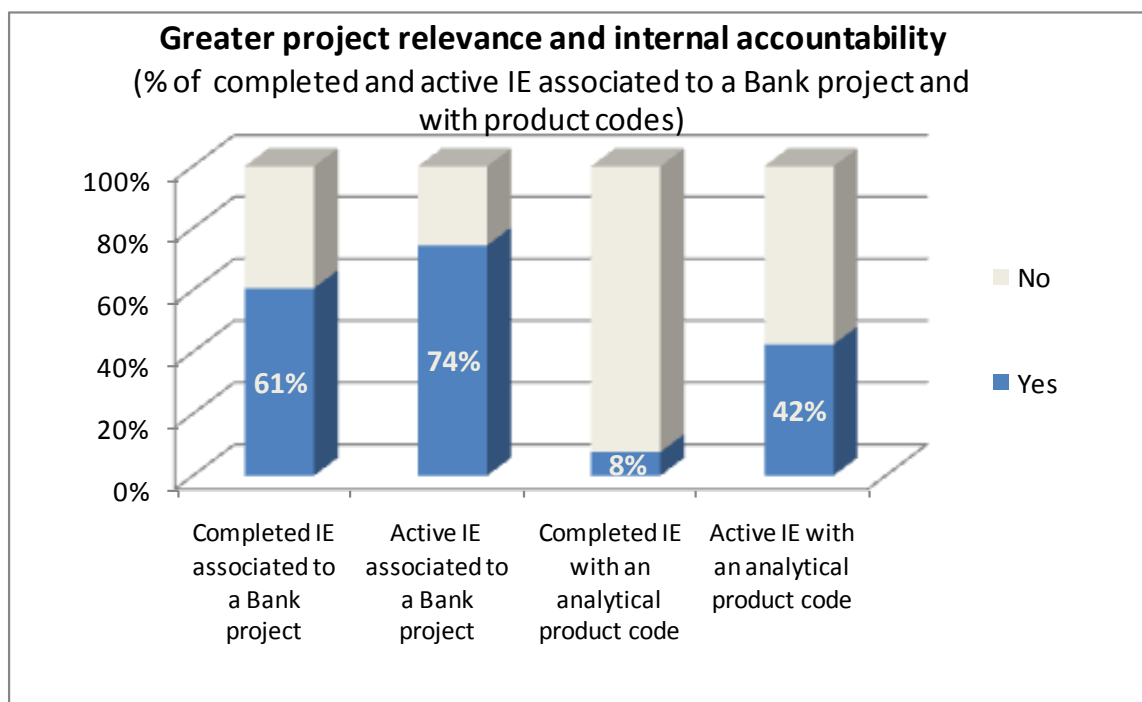


Figure 11. Quality Assurance

⁶ For example the DIME-FDP counts with high quality technical advisory group that includes researchers from the Bank, MIT, Yale, the London School of Economics, and the Harvard Business School.

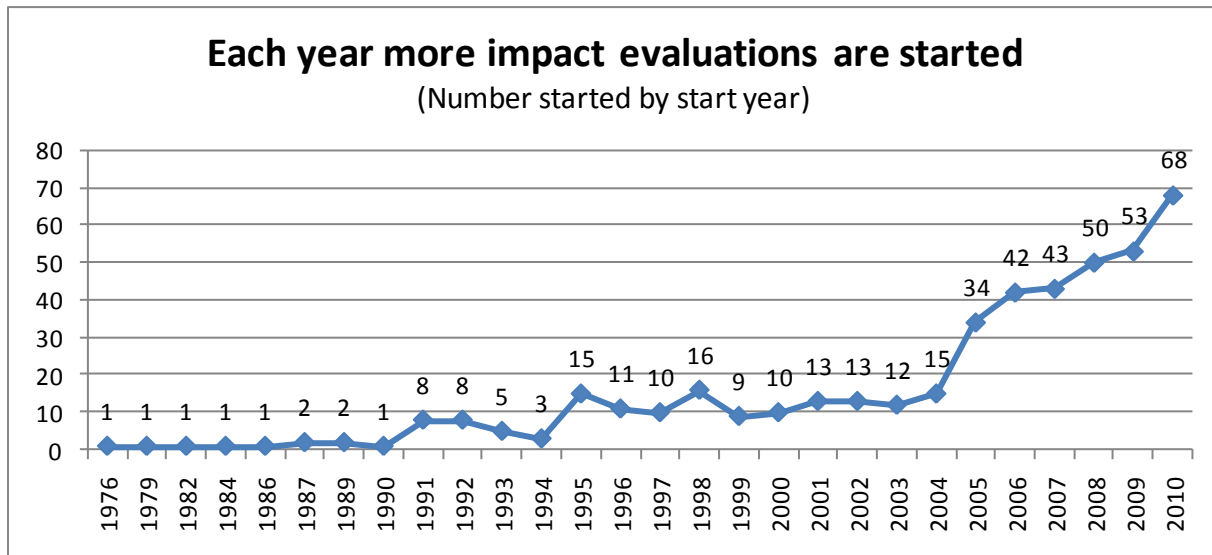


Figure 12. Number of IEs started by fiscal year, 1976 - 2010

Status of thematic programs. The number of evaluations experienced a significant expansion during 2009/10 with active impact evaluations growing from 190 in 2008 to 278 in 2010. Each year more evaluations are getting started (Figure 12), and most thematic areas are active (Figure 13). But not all areas are organized programmatically and not all evaluations are part of DIME programs. For example, the FPD evaluations that predate the 2009/10 period are mostly evaluations of IFC projects that are contracted and executed externally to the Bank with little client participation. A majority of all Bank IEs and 80 percent of IEs associated with Bank projects are supported by DIME structures. The objective is to increase the share of impact evaluations supported by DIME over time to ensure that best practice approaches are followed and that the evaluations can benefit from the community of practice and capacity building activities promoted by DIME.

The DIME programmatic model described in this paper was first tested by the Africa Program for Education Impact Evaluation (Legovini, Meky, and Tan 2007). It was then adopted by the Malaria Impact Evaluation Program (MIEP) (Friedman, Legovini, and Velenyi 2007), the Health Results-Base Financing IE program (HRBF) (Martinez 2007), the Africa Impact Evaluation Program on HIV/AIDS (AIM-AIDS) (Legovini, Lule, Bassole 2008), the Agricultural Adaptations and Rural Development IE program (AADAPT) in Africa, Latin America and South Asia (Goldstein et al. 2009), the Active Labor Market IE Program (Almeida 2009), the Gender Program (Buvinic 2010), the IE program in Finance and Private Sector (DIME-FPD) (Legovini, McKenzie, Stein 2010) and the Fragile States Program (DIEFS) (Legovini and Lierl 2009).

However, DIME is not the only model for implementing IE programs. Each network has a different history and approach:

1. HDN considers impact evaluation as a core area of responsibility of the Anchor and sector units in the regions have gone a long way in advancing impact evaluation. The Anchor rolled out its thematic “clusters” focusing on technical quality, instituting a competitive award system of (SIEF) financing, and catering to the evaluations

awarded SIEF grants. The clusters were smaller than the underlying programs and were not initially designed to conduct programmatic activities. Each SIEF cluster was assigned two part-time senior cluster leaders and no full time coordinator. The impact evaluation training program of HDN, which was aimed at expanding capacity for impact evaluation among governments and academics, was not designed to target existing clusters or cluster participants. HDN—a full DIME partner—is moving from the original model to ensure that the clusters are inclusive and not defined by the source of funding. Furthermore, HDN have started conducting programmatic activities targeted to cluster participants including training and dissemination. At least three HDN programs have adopted a full programmatic model.

2. In the case of access to infrastructure, the Sustainable Development Network (SDN) anchor provides light coordination, mainly to support financing and keep an eye on progress. SDN senior management communicated its intention to establish a core unit of impact evaluation specialists to launch three new programs (energy mitigation, forestry adaptations and water resource management) and support the existing program transition to an integrated model of implementation. This decision is yet to be operationalized. In the meantime, SDV decided to partner directly with the DIME Secretariat in the implementation of local governance reforms and fragile states.
3. The PREM Anchor already has a core unit to manage the program of institutional reform. The program will be formally launched during 2010 to start the implementation of programmatic activities. In addition, PREM gender adopted a gender mainstreaming approach and provided technical and financial resources to the AADAPT and DIME-FPD programs to rigorously identify gender effects in Bank operations. The approach appears to successfully integrate gender concerns into project design, management, monitoring and evaluation (Buvinic 2010).
4. The FPD network has enthusiastically embarked on a close collaboration with DIME. The DIME-FPD program, coordinated by DEC, was launched in February 2010 with high participation from senior management and sector teams from four regions.

Overall, the speed of adoption has been surprisingly high given the paradigm shift that the approach entails. In operational practice, this involves a search for empirical evidence to guide policy recommendations. For the Bank, this means a transition from knowledge to learning and an expanded need to support technical assistance activities.

Completed and active impact evaluations by DIME program (2010)

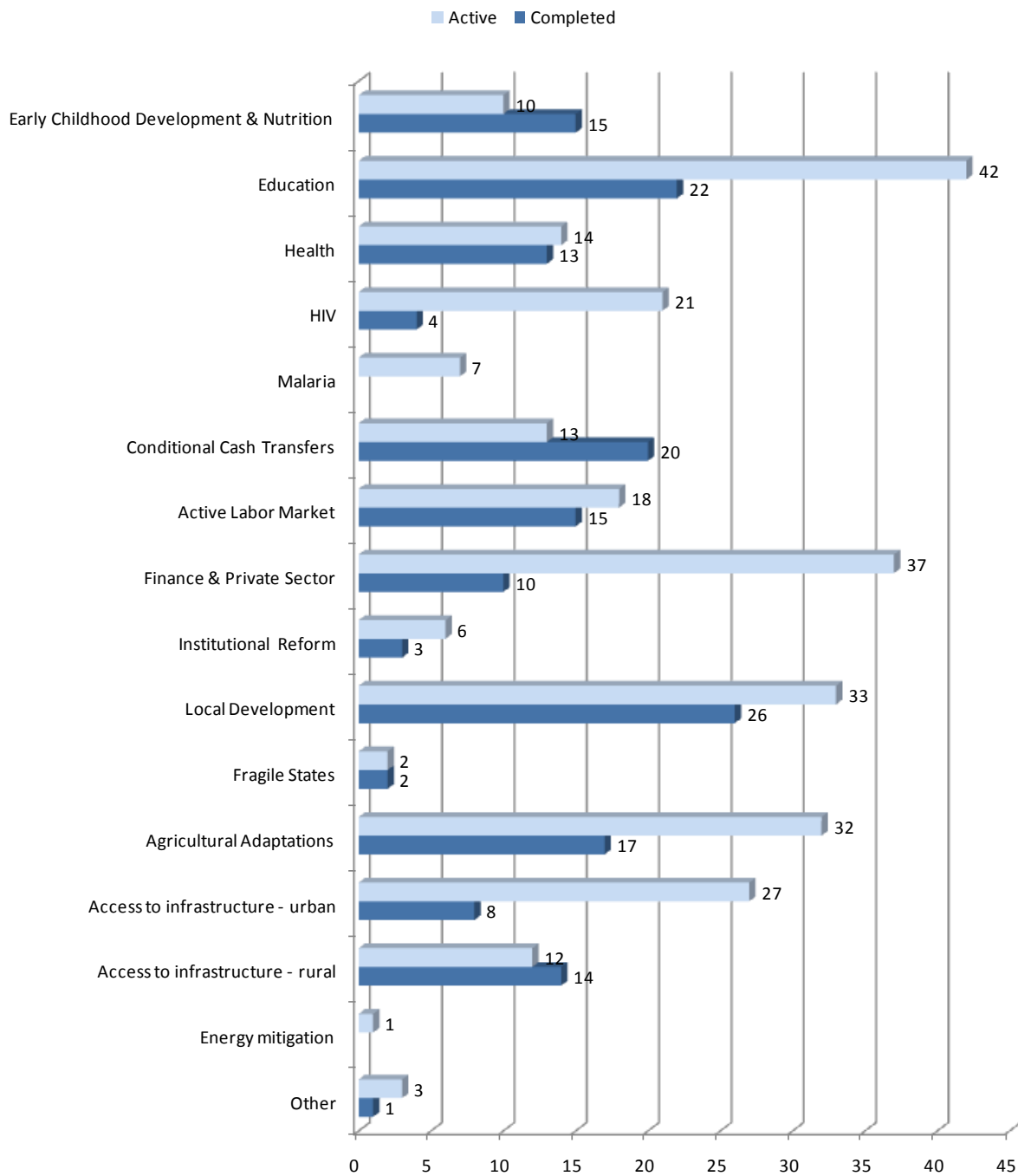


Figure 13. IE by thematic program

DIME Methods

“Experimental interference is of enormous importance, because without it you can never be sure that the correlation you observe has any causal significance.” -- Richard Dawkins, Professor for the Public Understanding of Science (retired), Oxford University

DIME conducts counterfactual analysis. Counterfactual analysis asks what would have happened without the intervention. It uses treatment-control comparisons to identify the causal effect of an intervention on outcomes separately from the effect of other time-varying factors. Treatment and control groups should have average observed and unobserved characteristics that are statistically identical before the intervention is implemented. The goal is for the two groups to have had the same outcome levels and trajectories in the absence of the intervention. This is to ensure that the intervention is the only difference between the two groups and that future differences between the two groups are uniquely caused by the intervention.

Post hoc ergo propter hoc.⁷ IE’s causal analysis is different from impact assessment and monitoring which involve before/after comparisons with no control group. Before/after comparisons describe changes in the treated group and not the reason why the changes occur. For the purpose of evidence-based policy, before/after comparisons can be misleading. For example, post-disaster improvements may be attributed to post-disaster relief, failing to account for concurrent factors such as victims’ own post-disaster recovery efforts. Monitoring and impact assessments should be used to track the efficiency of implementation, not the effectiveness of interventions.⁸ Impact evaluation is also different from poverty and social impact assessments that simulate the distributional effects of a program ex-ante, economic analyses that compare benefits with costs, and participatory assessments that rely on qualitative information gathered through interviews, beneficiary surveys, and stakeholder meetings (World Bank 2009a).

Experimental and non-experimental methods. Methods used for causal analysis are either experimental or non-experimental (Box 7). Experimental methods use ex-ante random assignment of the intervention to obtain comparable treatment and control groups. Random assignment produces statistically identical control and treatment samples in both observable and unobservable characteristics. Impact is measured as the ex-post mean difference in outcomes between treatment and control group. The resulting estimates deliver unbiased estimates of average impact. Most prospective evaluations (designed ex-ante) use experimental methods.

Non- or quasi-experimental methods mimic experimental design and use econometric analysis to estimate impact. They generally require more assumptions. Because most non-experimental methods cannot control for the effect of unobservable characteristics on program assignment, the estimates suffer from omitted (unobserved) variable bias. Non-experimental methods are used when randomization is not feasible and must be coupled with robustness checks to understand the direction of the bias. For example, retrospective evaluations, which are designed after an intervention has already taken place, and evaluations of bulky infrastructure use non-experimental methods.

⁷ “After this, therefore because of this” is a logical fallacy that reminds us that correlation is not causation.

⁸ Projects with an IE make use of *intensified* monitoring to ensure compliance with design.

A Glossary of common evaluation terms and impact evaluation methods

Impact: Effect of an intervention on any outcome

Causal inference: Finding the cause-effect relationship between intervention and outcome

Prospective evaluation: An evaluation designed before and evaluated during/after the intervention

Retrospective evaluation: An evaluation designed and evaluated after the intervention

Ex-ante evaluation: An evaluation carried out through simulations before the intervention

External evaluation: An evaluation designed and carried out by evaluators who are external to the implementing agency, generally for accountability purposes (measure results)

Internal evaluation: An evaluation designed and carried out by the implementing agency with the help of external evaluators, to guide implementation (managing for results) in addition to measuring results

Summative evaluation: An evaluation that measures or summarizes results

Formative evaluation: An evaluation that helps fine-tune a program for better results

Identification strategy: An analytical design to isolate the causal effect of an intervention on an outcome

Quantitative: Using numeric data

Qualitative: Using non-numeric observations

Interval validity: Soundness of the identification strategy for the sample included in the evaluation

External validity: Representativeness of the evaluation results for the population of interest

Experimental evaluation: Randomly assigning units of observation (households, schools, communities) to treatment or control to identify causal relationship between intervention and outcome. Assures balance in observed and unobserved characteristics and generates unbiased impact measures. Includes:

- Randomization: This gives all units in the study the same chance of being assigned to the treatment or control group. Provides unbiased impact estimates of average treatment effects.
- Random phase-in: Control units receive treatment later on. The approach cannot look at long-term effects except for populations that become ineligible later on (e.g. students graduating from school, retiring labor force).

- Random encouragement (a special case of IV): Experimental application of an instrument to cause exogenous variation in take-up. Used to estimate unbiased local treatment effect of the program on units responding to the encouragement.

Non- or quasi-experimental evaluation: Tool-kit of econometric methods that mimics random assignment. Assures balance in observed characteristics. Includes:

- Regression discontinuity (RDD): Compares participants and non-participants that are close to the eligibility criterion. Estimates unbiased local treatment effects, i.e. for the units of treatment that are close to the criterion. Often used in the presence of geographical, age, or other type of discontinuities that determine access to intervention benefits.
- Instrumental variable (IV): Use of a variable that determines exogenous variation in program take-up and is uncorrelated to program outcome to control for selection into the program. Estimates local treatment effect.
- Pipeline comparisons: For interventions that are rolled out sequentially along exogenous paths (e.g. roads); compares earlier units treated with units treated later.
- Difference-in-difference: Compares participants and non-participants before and after. The approach “differences out” ex-ante mean differences from ex-post mean differences to estimate average treatment effects. Assumes participants and non-participants evolve along the same trends. Bias of the estimate will depend on the validity of the assumption.
- Propensity score matching: Matches participants with observationally similar non-participants and compares outcomes for participants that have a match. Provides estimates of average treatment effect for observations that have a match. Bias is due to unobservables. Often combined with difference-in-difference to improve precision of the estimates.

DIME promotes prospective impact evaluation, an evaluative process planned before implementation (or expansion) of the intervention that measures results during and after implementation. Working prospectively has many advantages:

1. **Informs design.** Researchers that are brought in early contribute by sharing existing evidence and modifying the design of the intervention.
2. **Stimulates critical thinking.** The policy maker can set the agenda for the evaluation and pose the policy and operational questions that she finds more important to answer.
3. **Develops robust designs.** The analytical design can be developed to answer the questions of interest, and good opportunities found in the roll-out to improve the analytical validity of the evaluation.
4. **Helps plans for data collection.** Early engagement offers the opportunity to plan for needed data collection and align monitoring and evaluation data requirements.
5. **Affects decisions in real time.** Evaluations that accompany the interventions can provide feedback during implementation to improve results, not only measure results at the end. It can also be used to track the trajectory of results over time to help us understanding causal transmission, nonlinearities, reversals and sustainability patterns.

Prospective evaluation explains why the prevalence of experimental designs has risen in the active DIME portfolio of impact evaluations (74 percent) relative to completed evaluations (24 percent) (Figure 14). This is because random assignment has to be done ex-ante and because governments prefer the analytical simplicity and communicability of randomization results (mean differences) over the complexities of econometric analysis.

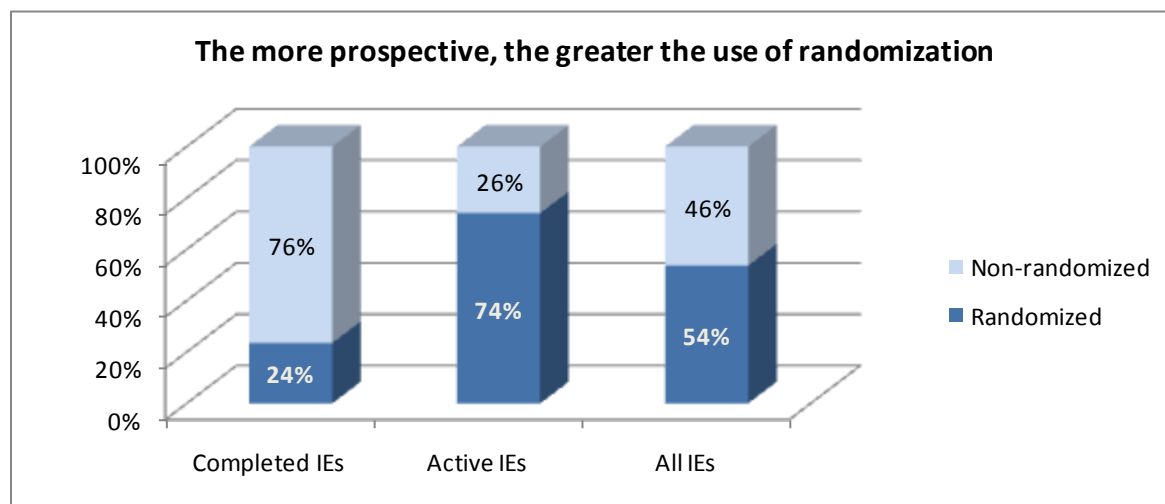


Figure 14. Randomized Methods

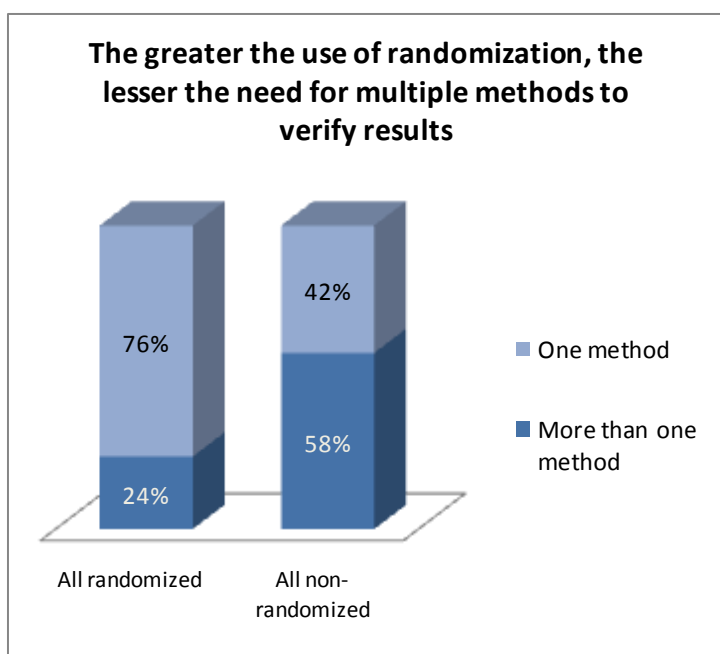
Some controversy surrounds the use of randomization. The issues raised include for whom the treatment effect is identified, whether the results are generalizable to other settings, and whether the causal mechanisms behind the workings of a policy can be understood (Deaton 2009, Ravallion 2008). These are important considerations that equally apply to the use of non-experimental methods and economic analysis in general (Banerjee and Duflo 2008). They should be addressed as part of a careful design and

interpretation of results. DIME works actively on these fronts to (i) understand how effects are distributed, for example, with gender mainstreaming; (ii) go beyond a simplistic black-box approach of *whether* something works or not to attempt to understand *how* something can be made to work and for whom, and (iii) by using the programmatic approach, model responses in different settings. Yet, as argued by Imbens (2009), when the question that one is interested in answering can be answered with a randomized design, there is little to gain and much to lose by not randomizing.

In practice, experimental and non-experimental methods are complementary and DIME uses the best feasible method to answer the questions of greatest policy relevance. Fifty percent of the evaluations use multiple methods within the same evaluative process. For example, the electrification of rural towns in Ethiopia is evaluated using geographical matching of towns and distance from the poles (non-experimental). The level of subsidy required to ensure high household connection to the grid is tested using town-wide lotteries (experimental) to ascertain that a small 10-15 percent subsidy is needed to reach the country target of 50 percent connection (Torero 2006). The irrigation scheme in Ethiopia is evaluated using a pipeline approach, reflecting the geographical roll-out of the scheme (non-experimental). Financial sustainability is tested by randomly assigning different sub-schemes to different repayment arrangements (experimental) (Kondylis 2008). Methods depend on the intervention type, implementation constraints and opportunities, and evaluation questions. Each case is *sui generis* and designed on a case-by-case basis. The use of randomization has reduced the need for multiple methods when used to check the robustness of findings (Figure 15).

Different methods can also be used to obtain answers to different policy questions. For a remedial learning program, in which students are coded as *A* for best, *B* for good, *C* for average, *D* for below average, and *F* for failing, the *F* students are targeted for the program. To evaluate the remedial learning we could use a randomized design to compare mean results for *F* students in the treatment and *F* students in the control. When the program considers whether to expand to *D* students, we would use a non-experimental method called regression discontinuity to compare the highest-performing *F* students in the program to the lowest performing *D* students outside the program to understand how the program may affect *D* students. The particular method partly depends on the question we seek to answer.

Figure 15. Use of multiple methods



Methodological constraints may affect what can be rigorously evaluated. Impact evaluation analysis relies on statistics and requires reasonably large numbers. This means that some policies may not be evaluated using counterfactual analysis. How large is large depends on the size of the effect expected, among other things. The greater the expected effect the smaller the number of observations required. This makes sense, since small effects are hard to distinguish from noise in the data. Reasonable sample size may go from a dozen units to many hundreds. In each case the sample size required to detect a minimum effect size will be determined using statistical methods. This explains why a policy that perfectly affects only one individual, one firm, one institution or one whole country at once cannot be evaluated using causal inference methods. There is no other individual, firm, institution or country that can reasonably serve as a good counterfactual. An example is the introduction of a new constitution or a new trade regime that instantaneously affects everyone at the same time.

The majority of policy and institutional reforms, however, can be rigorously evaluated. Most universal policies take time to roll out, affect different individuals differently, and, even when meant to provide universal access, may have imperfect (less than universal) take-up. Each case is assessed individually and opportunities for evaluation investigated. For example, the analysis of the effect of business registration reform in Mexico exploited the fact that the reform was introduced in different municipalities at different points in time. The reform increased employment by 2.8 percent (Bruhn 2008). Building on this experience, the state of Minas Gerais in Brazil is planning to use a randomized design to measure the impact of formalization on firm outcomes (Legovini et al. 2010). Similarly, De Janvry, McIntosh, and Sadoulet (2008) examine the impact of the introduction of a credit bureau in Guatemala. When a random sample of 5,000 microfinance borrowers received information about the credit bureau, repayment rates increased. Other examples include judicial reform in Senegal, trade facilitation in South Africa, and national fertilizer subsidies in Tanzania.

Qualitative complement quantitative methods. Qualitative methods such as classroom or other institutional observations, structured interviews and focus group discussions can be used to inform the design of the data collection instruments and deepen the understanding of the causal transmission mechanisms. These methods are critical for evaluating complex institutional changes for which there exists little ex-ante knowledge. For example, the experimental *quantitative* evaluation in India of a grassroots effort to energize citizens to participate in the village government is complemented by *qualitative* work. The latter is used to understand the dimensions of change: political and social dynamics, corruption, economic changes, and network affiliation. The month-to-month changes observed in the treatment and control areas take an in-depth look at the quality of the treatment and the changes it introduces (Rao 2010). Combining qualitative and quantitative methods capitalizes on their strengths and is different from traditional participatory assessments—widely used in Bank projects—that only hold discussions with participants at closing and should not be used to evaluate impact.

DIME Contribution to the Results Agenda

DIME is fully aligned with the Bank's results agenda. It introduces impact evaluation to enable the Bank to rigorously measure and improve results. This is its main contribution. Further, DIME makes a positive contribution to the quality of project data and M&E by complementing sector skills in the operations with Bank-delivered technical assistance specialized in data and analysis.

Data quality, availability and usage

World Bank projects have spent an estimated \$2 billion for M&E in the last 10 years,⁹ about a quarter of this in data collection. The funds are used to hire M&E consultants, collect baseline, midterm, and follow-up data for monitoring purposes, build monitoring systems, and prepare M&E reports. A recent OPCS review of 162 projects approved in fiscal 2009 reveals that 51% of projects had baseline statistics for all the main indicators in their results framework and 87% of project had summary statistics for at least some of their indicators.¹⁰

Much can be done to improve the quality and frequency of data when the process is informed by a data collection expert or researcher. Questionnaires designed for economic analysis are more useful than those designed solely to estimate summary statistics for the purpose of informing the results framework in the ISR. A field-based team helps increase data quality by strengthening local data collection capacity, supervising data collection and training enumerators. Use of consistent terms of reference ensures that consultants submit the raw data files and provide the information needed to catalogue data and make it available and usable.

A stock-taking of the M&E in the India portfolio reveals that the 34 surveys conducted in the fiscal 2004-2007 period were one-shot exercises. In no case did the surveys follow a baseline/follow-up strategy. They are therefore of limited use for monitoring purposes. There was almost no staff time allocated to ensuring the quality of the data collected. Only a combined total of 12 staff weeks were used for the contracting, supervising, and report writing of 15 surveys. Use of data was similarly low. Only six working papers were generated using the 34 datasets. On average, each dataset was used to produce 1.17 outputs (presentations, notes, papers). The data cost per output varied between \$23,000 and \$90,000. Finally, data is currently not available. Of the 34 datasets, only 4 are available to the Bank and 7 to the task teams (Shah and Fiszbein 2009).

The Data Group produced a preliminary estimate of the size of the investment in project data and the low returns from it. In the fiscal 2000-2009 period, World Bank lending and technical assistance projects spent \$419 million on data collection. Project preparation grants spent US\$37 million in the fiscal 2000-2004 period. However, only an estimated 5 percent of those datasets are available in any Bank data repository (Mistiaen 2010).

⁹ The Bank does not record contracts by purpose. The figure is a lower bound estimate using a 125 key-word search across FY00-FY09 project contracts. The review of these contracts suggests they are for data collection and consultancies for M&E. A more accurate estimation requires drawing a random sample of projects, reviewing content and amount of all contracts, and expanding the estimate to the whole Bank portfolio.

¹⁰ The definition of baseline in this paper is different from the IDA15 definition. The latter requires summary statistics for the development outcome indicators in the ISR, not a full baseline data for analytical purposes.

DIME works in collaboration with project teams to address some of these problems, especially the issues of data quality and data access. Research teams help projects develop a baseline and follow-up data strategy for data collection, design the questionnaires and samples, and train enumerators and supervise data collection. DIME also worked to develop the tools to upload and document data for other people to use. Furthermore, all data is analyzed. Of the 310 projects that have an impact evaluation, so far 176 have at least one round of survey data and an estimated 126 have at least two rounds of data. All are expected to have at least two rounds of data by completion. While not all projects may need survey data, a consistent strategy for data collection is important to increase the quality of project M&E, the returns to data collection and the ability to inform the policy process.

M&E systems

Good information systems can be used for regular and periodic impact evaluation. A good information system for the purpose of evaluation has reliable data quality on indicators of interest and includes data on both treatment and control groups. In the cited example of the evaluation of HIV prevention programs in Senegal, all information came from routine data from health centers and districts. Analytical structure alone—in that case a random assignment of health districts to different interventions—enabled the use of routine information for the purpose of rigorous impact evaluation. Availability of information systems and adequate design features combined with analytical structure can therefore significantly increase the ability of agencies to conduct routine impact evaluation at very low cost.

However, seldom are minimum conditions available for using information systems for impact evaluation. The main issues are data coverage of control individuals or areas and the structured roll-out of policy changes, interventions or innovations *ex ante*. Data quality is also often at issue. When impact evaluation teams work hand in hand with program teams and system developers, outcomes can be improved. An excellent example is from India's Health Insurance Scheme. The Bank research team invested their time and resources to develop in-house the monitoring systems and complementary survey data systems that ensure the flow of high-frequency and reliable data into real-time decision-making (Box 8). In Eritrea, a USAID malaria impact evaluation research team organized all administrative and outcome information in selected regions of the country. The Ministry of Health took advantage of this initial effort, expanded it to the other regions and created the most comprehensive and regularly updated national system to track malaria activities and outcomes (Graves 2004). In South Africa, the Department of Education provided the Bank access to existing and separate data systems. The research team worked to assemble the data to evaluate the impact of Dinaledi, their science and technology program. The work was fraught with difficulties because schools did not count with a single identifier and records had to be matched one by one. The data merge took two months of full-time work. When in 2008/2009 the Department of Education became aware of the problem, it introduced a single numeric identifier for every school in the country. This will facilitate the regular monitoring and evaluating of school programs from now onwards (Blum, Krishnan, and Legovini 2010).

Attribution

The final and perhaps most important contribution of impact evaluation to the results agenda is to enable the attribution of results to the intervention in question. Before/after analysis can only tell us that things changed but cannot identify the reason why.

Counterfactual analysis can point to the exact cause-effect relationship and tell us how much of the change is caused by the intervention. This can inform the decision to scale up or down, increase or decrease financing, or modify the intervention to ensure better results.

In essence, what differentiates monitoring from impact evaluation is analytical structure and not the type of data that is required. Seldom does analytical structure affect costs. The question regarding what should be the optimal scale in the use of impact evaluation in Bank projects is therefore not one of impact evaluation-specific costs, but of the costs of improving data systems and building awareness and capacity internally and externally.

Overall, the combination of better data with an analytical strategy to understand the cause-effect relationships helps improve the quality of the project, inform the Mid-Term Reviews and Implementation Completion Reports, and strengthen OPCS' ability to track progress and report results Bank-wide. IEG can also greatly gain from DIME's work through the availability of good quality data at baseline and endline. The analytical structure of projects with impact evaluation can help IEG assess project implementation quality and effectiveness with greater rigor.

Box 8. Getting M&E right: India's health insurance scheme

In 2008, India launched its flagship health insurance scheme for the poor--the Rashtriya Swasthya Bima Yojana or RSBY. Under the scheme, eligible households enroll and receive a smart card which they can use in a network facility—public or private—for inpatient care for up to Rs. 30,000 a year. Insurance companies bid for project districts and the government pays the winning companies the premium for the households enrolled under the scheme. The scheme incorporates the latest thinking to address the problem of health provision in low-income countries: it empowers households by giving them choice; it incentivizes insurance companies by conditioning payment on enrollment; and it induces hospitals to provide better care through increased competition. The scheme uses biometric smart cards to manage fraud and collect data.

From the scheme's inception, the World Bank has been closely involved in providing monitoring and evaluation guidance for this important scheme. The Bank, together with a local IT firm (Pratyaksh), designed a data submission system for eligible households—the backbone of the scheme. The Bank then helped finance the development of the enrollment and transaction software and designed the Monitoring and Information System to capture data through the smart cards and feed it back to policy makers and enable real-time decision making. Through specialized surveys of households, hospitals and scheme beneficiaries, the Bank developed a deeper understanding of how the scheme is working. Institutionalizing these surveys will enable close monitoring of the functioning of the scheme. Finally, the Bank also piloted interventions to help understand how the scheme can be improved. Two of these examine the impact of information on enrollment and of health camps on hospitalization. Soon, the Bank with other research collaborators hopes to release a set of findings on different aspects of the scheme in its first two years to enable discussion and debate.

The result: (i) the largest health administrative data system in the world updated in real time by electronic data flows from smart cards, and (ii) evaluations grounded in the functioning of the health insurance scheme which pilot results that affect real-time policy decisions.

Of the several flagship programs of the Government of India, this is the first where the monitoring and evaluation is directly tied into the functioning of the program. The Government of India recognized the key role of the World Bank in the design of M&E systems that take into account the administrative situation, the incentives of different players in the scheme, the downstream analysis that can be performed and the issues that will arise in the future.

Early on, team members--drawn equally from operations and research---recognized that harnessing these unique advantages required a full "hands-on" approach. The team would fully design and implement various monitoring and information systems and once they understood the benefits and costs of different approaches, they would be in a better position to "contract out" these systems to external firms. This has not been easy. The team has operated on a shoestring "preparation" budget and it has been very difficult to obtain funding for this venture. How can things be changed so that M&E in flagship schemes like these, with repercussions for countries around the world, can be appropriately funded? One approach would be to institute a "global fund" that identifies such critical programs and provides them the support required to achieve their full learning potential.

Jishnu Das
Senior Economist, DEC Research Group

DIME Human and Financial Resources

Skills

DIME is a decentralized effort with a small number of staff centered in the DIME Secretariat and others scattered across 22 separate units with varying degrees of impact evaluation skills and time allocated to the task. There are an estimated 79 people that are doing impact evaluation and have impact evaluation skills in the Bank. One third of them are in DEC, one third in Africa and HDN combined, and the remaining third in all the other networks and regions combined (Figure 16). The number in DEC reflects the current engagement of the DIME Secretariat and the Research Group in operational research. The large numbers in Africa and HDN reflect the management of the Africa Impact Evaluation Initiative and the SIEF.

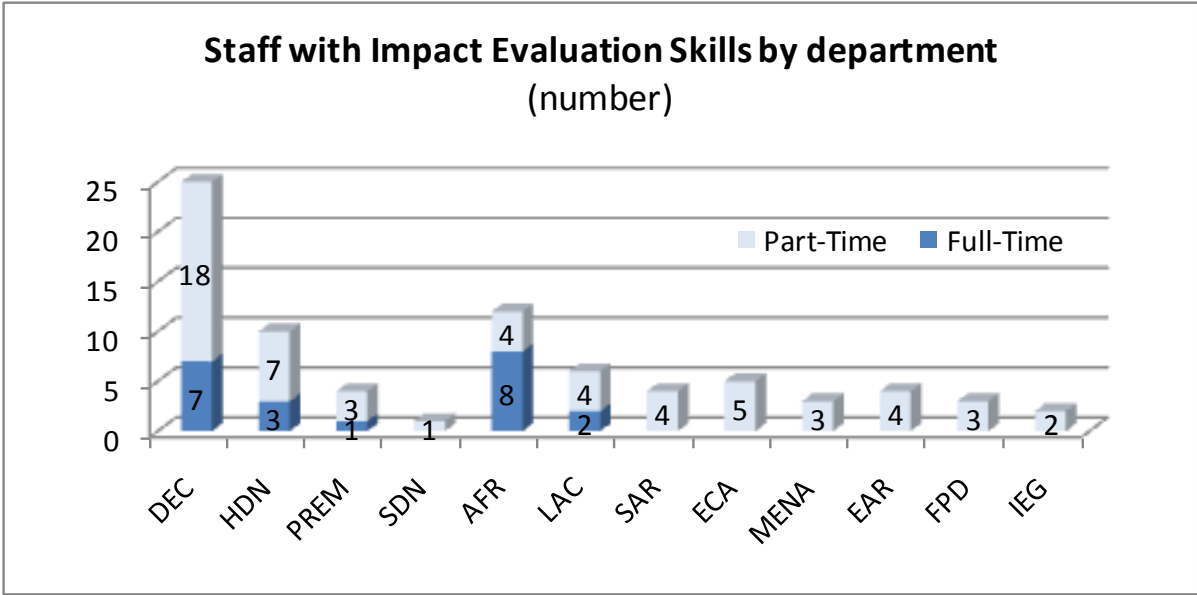


Figure 16. Staff with impact evaluation skills

Of the 79 staff identified, 21 work on impact evaluation full time, mainly in DEC (7) and in Africa (8). They manage the impact evaluation portfolio with an average oversight of 15 impact evaluation products per staff. Two-thirds of them are F level staff or equivalent (Figure 17). About half of these full-time staff are on consultant or co-terminous contracts (11, Figure 18). The remaining 58 work on impact evaluation products part of the time. Most of them are permanent staff (51).

As the DIME initiative becomes mainstreamed into operational work, strategic IE staffing in sector units should increase.

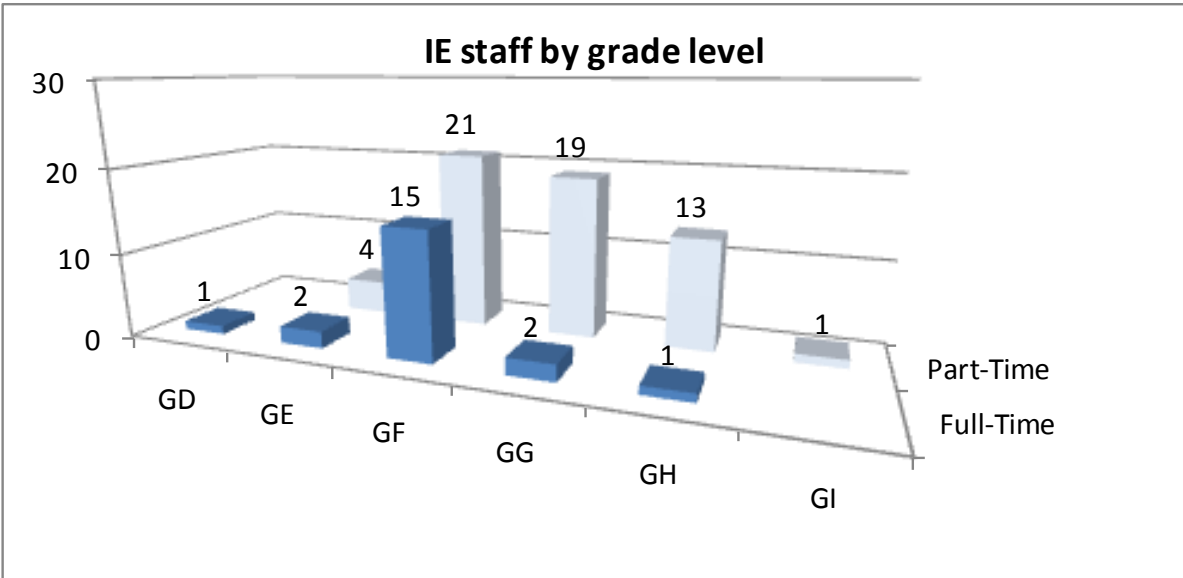


Figure 17. IE staff by level

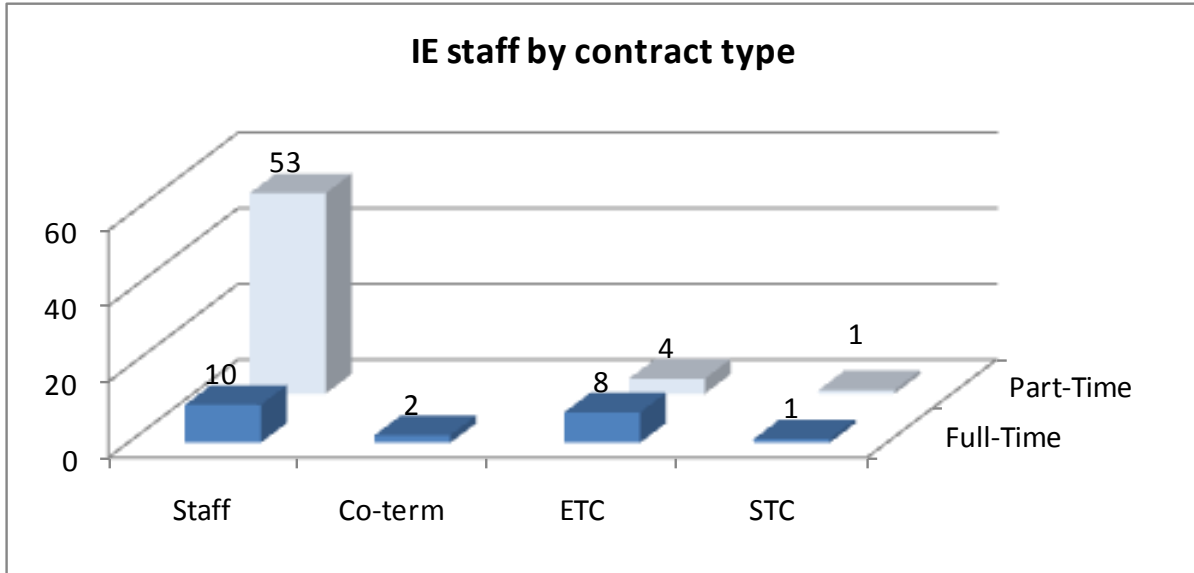


Figure 18. IE staff by contract type

Costs and financing

In order to promote learning and positive externalities, the Bank shares the costs of impact evaluation with clients. The Bank uses Bank internal funds and trust funds.¹¹ The client uses project financing. Full accounting of the costs is an arduous process both because the financing of impact evaluation is complex and shared between multiple sources and because many impact evaluations do not count with a separate product code. To estimate the costs of conducting impact evaluation at the Bank, two separate exercises were undertaken: (i) to track product codes that do impact evaluation; and (ii) to track project financing of impact evaluation and contracts for impact evaluation.

¹¹ Supporting donor trust funds include, but are not limited to, BNPP, BPRP, LPRP, DFID, RSB, EPDF-FTI, GAP, HRBF, IDF, KCP, SIEF, TFESSD, and UNAIDS/UBW TF.

Bank financing

A review of all IE and other analytical (AAA) product codes was used to verify and identify impact evaluation activities. In addition to IE codes, research (RF), knowledge (KP), and technical assistance (TA) codes are sometimes used to conduct impact evaluation. The exercise found 153 active and 15 closed codes for specific IE products and 38 active codes for IE coordination. This leaves 45 percent of active impact evaluations without an analytical budget code. Therefore, the Bank-executed spending computed from SAP is widely underestimated.

During fiscal 2000-2004, the Bank spent \$160,000 a year on impact evaluation in total. After DIME was established, spending rose to \$4 million a year in fiscal 2005-2008, evenly split between internal and trust fund resources. Since DIME was re-launched as a Bank-wide initiative and with the maturing of the IE portfolio, spending tripled to \$13-14 million a year, mostly financed by trust funds (Figure 19). The proportion financed by trust funds rose to 80 percent of the total.¹² Furthermore, internal resources fell in absolute terms from \$3.7 million in fiscal 2009 to an estimated \$2.7 million in fiscal 2010. Of this, \$0.7 million (or 25 percent) came from the DIME Research Support Budget, the allocation for which ends in June 2010. Internal funding for IE could fall in fiscal 2011 to half the level of fiscal 2009. A lack of sufficient internal resources could severely undermine Bank-wide efforts to mainstream IE into Bank's operations and to leverage trust fund resources.

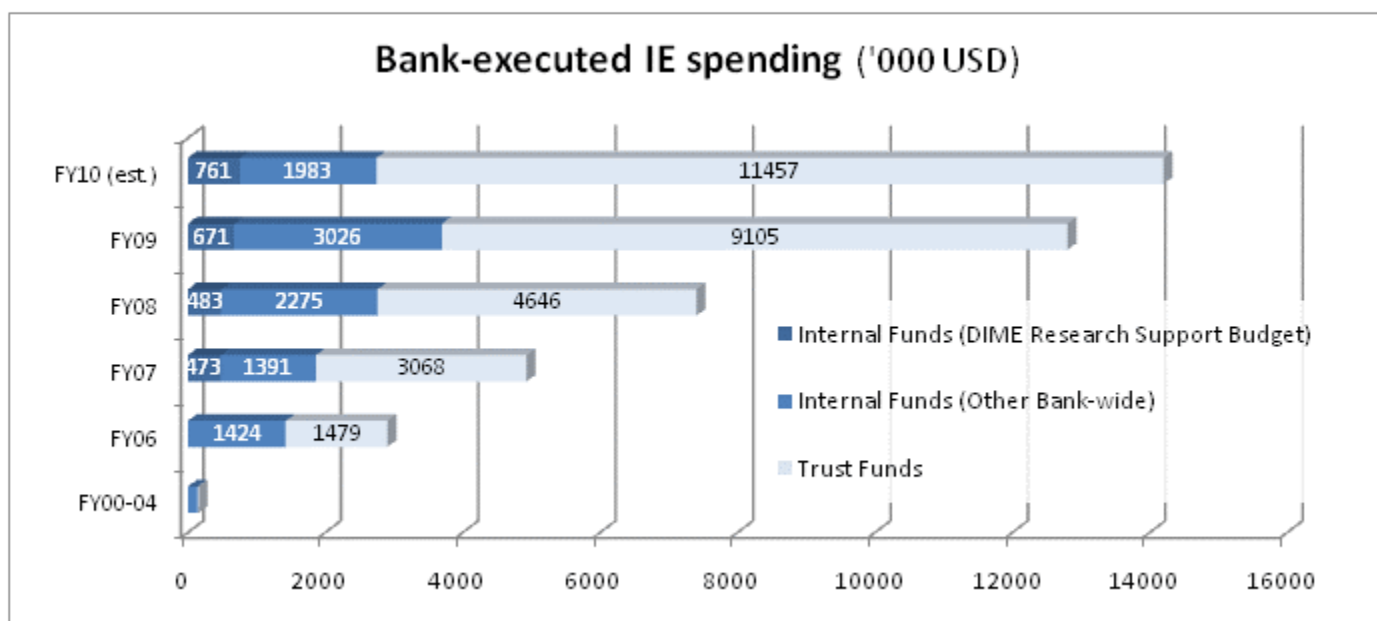


Figure 19. Bank-executed IE spending, FY 2000-2010

¹² During January 2009 and June 2010, the IE programs coordinated by the DIME Secretariat alone raised \$13 million in grants. Most of these grants are used for country-specific products. Twelve percent is devoted to capacity development activities, 8% to coordination and the rest for country specific products (80%).

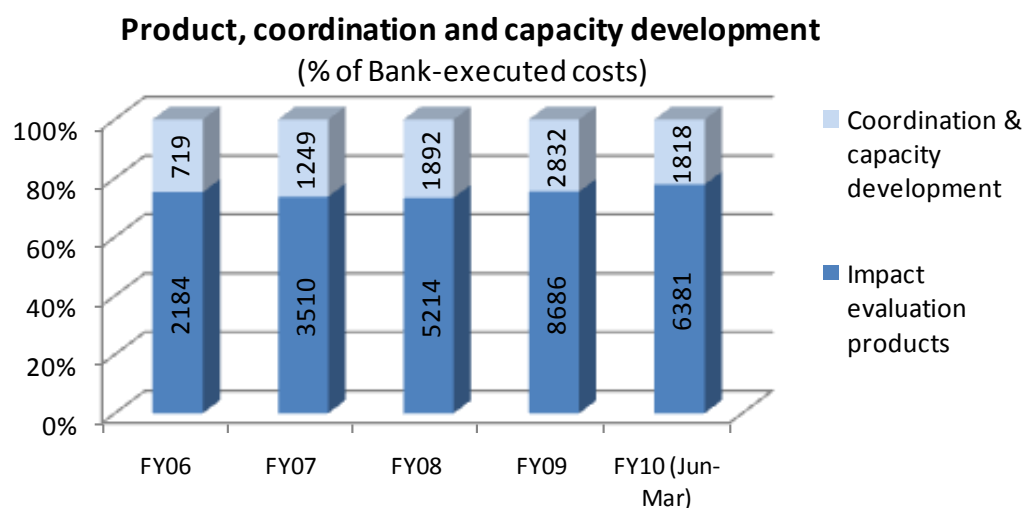


Figure 20. Spending by product type, FY 2006 - 2010

The IE spending contributes to country-specific products and pays for the cost of program coordination and capacity development. In total, one-fourth of Bank-executed spending goes into coordination and capacity development and three-fourths into specific analytical products (Figure 20). For example, in its first three years of operation, the APEIE program will have supported 14 countries and spent \$2.4 million (or 29 percent) in coordination and cross-country capacity development activities, and \$5.8 million (or 71 percent) on impact evaluation products and in-country capacity development (Table 1).

Table 1. Africa Program for Education Impact Evaluation budget (14 countries), FY 2007-2010

	WB project funding	WB BB funding	Donor funding	Total	% of total
Program management & capacity building	\$ -	\$ 162,600	\$ 2,205,896	\$ 2,368,496	29%
Program management		\$ 64,000	\$ 778,101	\$ 842,101	10%
Advisory board		\$ -	\$ 129,000	\$ 129,000	2%
Cross-country capacity development		\$ 98,600	\$ 1,248,795	\$ 1,347,395	17%
IE preparatory activities		\$ -	\$ 50,000	\$ 50,000	1%
Impact evaluation products	\$ 735,000	\$ 158,000	\$ 4,859,860	\$ 5,752,860	71%
% of total	13%	3%	84%	100%	
Product cost per IE	\$ 52,500	\$ 11,286	\$ 347,133	\$ 410,919	
Total cost	\$ 735,000	\$ 320,600	\$ 7,065,756	\$ 8,121,356	100%
% of total	9%	4%	87%	100%	
Total cost per country	\$ 52,500	\$ 22,900	\$ 504,697	\$ 580,097	

A large share of spending for capacity development goes into the cross-country workshops of the Africa Impact Evaluation (AIM), DIME Secretariat, and HDN (Table 2). In fiscal 2005-2010, 3,100 people or 450 teams were trained on impact evaluation and exposed to international evidence at a cost of \$1,400 per person or \$9,900 per team. The efficiency rate for AIM and DIME Workshops was 86 percent: 144 of the 167 teams which participated are implementing their impact evaluation. For each evaluation going forward one or more workshops were conducted in-country at least doubling the number of people trained in impact evaluation.

Table 2. Cross-country impact evaluation workshops

DIME Program	Region	Location	Date	FY	Cost USD '000	Finance	Total participants	Bank staff	Government	Other	Teams participating	Teams doing IE	Ratio	Responsibility
Cross-sector	AFR	Kenya	Jun-05	2005	50	PADI	43	10	33	0	8	6	75%	AIM
Cross-sector	AFR	South Africa	Jun-06	2006	90	South Africa	219	11	177	31	8	4	50%	AIM
Malaria	AFR	South Africa	Apr-07	2007	145		45	10	35	0	4	3	75%	AIM
Educ.	AFR	Nigeria	Jun-07	2007	242	EPDF/EFI-FTI	60	12	48	0	12	12	100%	AIM
HD	MENA	Egypt	Jan-08	2008	175	SIEF	160	21	106	33	33	NA		HDN
Malaria & HIV	AFR	Eritrea	Feb-08	2008	200		95	28	67	0	16	8	50%	AIM
HD	LAC	Nicaragua	Mar-08	2008	125	SIEF	104	19	85			NA		HDN
HD	NA	Spain	Jun-08	2008	225	SIEF	196	18	104	74	74	NA		HDN
Educ. & HIV	AFR	Senegal	Dec-08	2009	332	EPDF/EFI-FTI	67	12	52	3	13	13	100%	AIM
HD	EAR	Philippines	Dec-08	2009	150	SIEF	137	18	97	22	22	NA		HDN
HD	LAC	Peru	Jan-09	2009	125	SIEF	182	16	138	28	28	NA		HDN
HIV	AFR	South Africa	Mar-09	2009	111	UBW	77	13	55	9	9	8	89%	DIME Secr/AIM
HD	MENA	Jordan	Mar-09	2009	150	SIEF	206	26	164	16	16	NA		HDN
Agric. & Loc.Dev. Labor Markets	AFR	Ethiopia	Apr-09	2009	210	GAP/BPRP	114	32	61	21	20	19	95%	DIME Secr/AIM
							35						100%	HDN
	LAC	USA	May-09	2009	25	BNPP		15	20		4	4		
HD	EAR	China	Jul-09	2009	250	SIEF	212	31	153	28	28	NA		HDN
HD	ECA	Bosnia	Sep-09	2010	150	SIEF	115	20	85	10	10	NA		HDN
Agric. & Loc.Dev.	LAC	Brazil	Nov-09	2010	120	GAP	100	22	64	14	13	9	69%	DIME Secr
Agric. & Loc.Dev.	SAR	India	Dec-09	2010	180	GAP/TFESSD	67	20	41	6	12	6	50%	DIME Secr
Health	AFR	South Africa	Dec-09	2010	125	SIEF	108	18	70	20	20	11	55%	HDN/AIM
HIV	AFR	South Africa	Jan-10	2010	NA	Gates	101	3	45	53	6	NA		GATES/DIME
FPD	Global	Senegal	Feb-10	2010	215	GAP	100	26	66	8	16	16	100%	DIME Secr
HD	SAR	Nepal	Feb-10	2010	125	SIEF	118	28	67	23	23	NA		HDN
HD	LAC	Brazil	Apr-10	2010	200	SIEF	150	30	100	20	12	NA		HDN
Educ.	AFR	Ghana	May-10	2010	392	EPDF/EFI-FTI	83	22	54	7	12	12	100%	DIME/AIM
Social Protect.	AFR	Ghana	May-10	2010	150	SIEF	125	25	90	10	12	N/A		HDN
Fragile states	Global	Dubai	Jun-10	2010	150	TFESSD/BNPP	100	20	65	15	13	13	100%	DIME Secr
TOTAL					4,412		3,119	526	2,142	451	444	NA	NA	
Efficiency rate for DIME/AIM workshops for which data is available (excl HDN)											167	144	86%	
Average cost ('000)							1.414 per person			9.936 per team				

Client financing

Clients pay for a large portion of impact evaluation costs, usually through project financing. The figures are available for the programs coordinated by the DIME Secretariat (Table 3). In most of these programs, the DIME Secretariat explicitly requires countries to allocate funds for data collection to access DIME-provided advisory services. As a result, client governments pay for about half of impact evaluation product costs, mostly for data collection. Government staff costs are not included. The justifications for Bank (and trust fund) financing are the learning and externalities involved in participating.

Table 3. Cost structure of impact evaluation products
(USD, DIME Secretariat-coordinated programs)¹³

	<i>Governments (Project financing)</i>	<i>WB Budget</i>	<i>Donor Funding</i>	<i>Total</i>
APEIE Africa Program for Education Impact Evaluation (14 IE, FY07-present)				
Total	735,000	158,000	4,859,860	5,752,860
% of total	14%	3%	83%	100%
Average per IE	59,357	11,286	347,133	417,776
MIEP Malaria Impact Evaluation Program (7 IE, FY08-present)				
Total	3,280,000	209,000	4,293,650	7,782,650
% of total	42%	3%	55%	100%
Average per IE	468,571	29,857	613,379	1,111,807
AADAPT-AFR Agricultural Adaptations Impact Evaluation Program in Africa (14 IE, FY09-present)				
Total	4,658,000	490,000	3,309,000	8,457,000
% of total	55%	6%	39%	100%
Average per IE	332,714	35,000	236,357	604,071
DIME HIV/AIDS Impact Evaluation Initiative (5 IEs ongoing + 5 in preparation, FY09-present)				
Total cost (FY09-10)	790,000	270,000	1,257,000	2,407,000
% of total	37%	11%	52%	100%
Average per IE	88,000	27,000	125,700	240,700
DIME FPD - Impact Evaluation in Finance and Private Sector Development (18 IE, FY10-present)				
Total	3,600,000	360,000	2,160,000	6,120,000
% of total	59%	6%	35%	100%
Average per IE	200,000	20,000	120,000	437,143
Total APEIE, MIEP, AADAPT-AFR, DIME HIV/AIDS, DIME FPD				
Total	13,063,000	1,487,000	15,879,510	30,519,510
% of total	51%	10%	38%	100%
Average per IE	225,224	25,638	273,785	526,198

Average cost of impact evaluation products

The costs of impact evaluation are small relative to the costs of the interventions they evaluate. During fiscal 2006-2010, actual Bank internal spending per impact evaluation

¹³ Data on the ADAPT LAC and SAR was not yet available at the time of the drafting of the paper.

product was \$54,000. This leveraged \$123,000 in grant funding and a quarter to half a million in government funding.¹⁴ Total Bank-executed spending represents on average 0.2 percent of an average loan. Total impact evaluation costs, including government contributions, average 0.6 percent of project cost, including data used for both monitoring and impact evaluation. For the programs coordinated by the DIME Secretariat, average cost per impact evaluation is \$526,198 divided between the Bank's internal funds (\$25,638), donors (\$273,785), and government (\$225,224). Relative to the size of the government programs that impact evaluations are meant to inform, costs are small.

Cost items include analytical services, IE management and data collection. Data collection is the largest cost item representing half or more of IE costs. Data costs vary across sectors and types of intervention, the level at which data is collected (e.g., household versus school), what data must be collected (e.g., biomarkers are expensive), and the frequency of an outcome (e.g., rarer events like deaths require larger samples). When the evaluation seeks to estimate effects on different populations and spillover effects, samples and costs increase greatly. Data costs are for both evaluation and monitoring purposes and tend to overstate the costs of evaluation. However, impact evaluation is scrupulous about executing the baseline-follow up strategy, and therefore surely increases project data costs relative to a project without impact evaluation.

The costs of impact evaluation costs and sometimes the cost of an intervention can be easily repaid with the small increases in project effectiveness, or by identifying costly and ineffective programs. This is the case, for example, with report card evaluations, the cost for which is covered by intervention activities.¹⁵ For example, Andrabi, Das, and Khwaja (2008) find that the welfare gains from implementing the impact evaluation of a report card system in the market for public and private education in Pakistan outstrip the costs. The provision of information increased learning by 0.10 standard deviations and decreased private school fees by 18 percent. The cost of implementing the intervention and evaluation was similar to the drop in school fees alone. When evaluations are based on administrative data, as was the case with the HIV evaluation in Senegal, the learning is done at almost no cost. In Senegal, \$50,000 in analytical services compared favorably to finding out how to double the number of HIV positive individuals getting tested.

Conclusions

DIME can report strong progress in advancing impact evaluation in the Bank and among clients. The initiative is aligned with the Bank's results agenda and helping to rigorously measure results and improve the quality and availability of project data. It is contributing to the Bank's knowledge agenda by generating knowledge in critical development areas. DIME is engaged in the process of improving client operations supported by the Bank by supplementing the skills needed for more evidence-based policy making. To do so, DIME is helping to move development thinking from prescriptive notions of what might work to the specific evidence needed to know how to make it work. The transformation, however, has only just begun.

¹⁴ Source: SAP actuals.

¹⁵ Report cards interventions collect data and disseminate findings to elicit demand and supply responses. By covering the costs of data collection and analysis, the intervention covers most evaluation costs.

References

Almeida, Rita. 2009. *Active Labor Market IE Program*. Concept Note. Washington, DC: World Bank.

Andrabi, Tahir, Jishnu Das, and Asim Ijaz Khwaja. 2008. *Report Cards: The Impact of Providing School and Child Test-scores on Educational Markets*.

http://www.hks.harvard.edu/fs/akhwaja/papers/RC_14Nov08.pdf.

Andrabi, Tahir, Jishnu Das, Asim Ijaz Khwaja, Tara Vishwanath and Tristan Zajonc. 2007. *Learning and Educational Achievements in Punjab Schools: Insights to Inform the Educational Policy Debate*. World Bank.

Arcand, Jean-Louis, Cheikhou Sakho, and Natascha Wagner. 2010. *HIV/AIDS Sensitization, Social Mobilization and Peer-Mentoring: Evidence from a Randomized Experiment in Senegal*. Unpublished manuscript. Washington, DC: World Bank.

Bamberger, Michael, Vijayendra Rao, and Michael Woolcock. 2010 (forthcoming). "Using Mixed Methods in Monitoring and Evaluation: Experiences from International Development." In *Handbook of Mixed Methods Research*, edited by C. Teddlie and A. Tashakkori. Sage Publishers.

Banerjee, Abhijit V., and Esther Duflo. 2008. *The Experimental Approach to Development Economics*. Cambridge: MIT. <http://econ-www.mit.edu/files/3159>.

Beltran, Maria Isabel, and Chris Reinstadtler. 2010. *Making the Grade*. Draft report. Washington, DC: World Bank.

Blum, Jurgen, Nandini Krishnan, and Arianna Legovini. 2010. *Doing the Math on a Math & Science Program: South Africa's Dinaledi*. Unpublished draft. Washington, DC: World Bank.

Bruhn, Miriam. 2008. *License to Sell: The Effect of Business Registration Reform on Entrepreneurial Activity in Mexico*. World Bank Policy Research Working Paper Series, #4538. Washington, DC: World Bank.

Bruns, Barbara, Deon Filmer, Harry Patrino. 2007. *Making Schools Accountable: What Works in the Developing World?* Concept Note. Washington, DC: World Bank.

Buvinic, Mayra. 2010. "DIME-Gender Action Plan partnership." World Bank presentation at the ARD Training - SDN Week, January 27, 2010.

Deaton, Angus. 2009. "Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development." In *Proceedings of the British Academy*, 2008 Lectures, Vol. 162. Oxford University Press.

Fiszbein, Ariel, and Norbert Schady. 2009. *Conditional cash transfers: reducing present and future poverty*. World Bank Policy Research Report. Washington, DC: World Bank.

Friedman, Jed, Arianna Legovini and Edit Velenyi. 2007. *SIEF Cluster Proposal for Malaria*. Washington, DC: World Bank.

Goldstein, Markus, Florence Kondylis, Nandini Krishnan, and Arianna Legovini. 2009. *Agricultural Adaptations and Rural Development IE program (AADAPT)*. Concept Note. Washington, DC: World Bank.

Graves, Patricia. 2004. *Eritrea: Malaria Surveillance, Epidemic Preparedness, and Control Program Strengthening*. Activity Report 144, EHP Project 26568/E.X.ER4.PUBS. Washington, DC: Office of Health, Infectious Diseases and Nutrition Bureau for Global Health, U.S. Agency for International Development.

Imbens, Guido. 2009. *Better LATE than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)*. Processed. Cambridge: Harvard University.

De Janvry, Alain, Craig McIntosh, and Elisabeth Sadoulet. 2008. *The Supply- and Demand-side Impacts of Credit Market Information*. Working Paper. San Diego: University of California-San Diego.

Kondylis, Florence. 2008. *Designing an impact evaluation for the Ethiopia Irrigation and Drainage Project*. IE Concept Note. Washington, DC: World Bank.

Legovini, Arianna, David McKenzie, Miriam Bruhn, Priscila Malaguti, and Joan Luis Soares. 2010. "Reducing informality among firms in Minas Gerais, Brazil". IE Concept Note (unpublished). Washington, DC: World Bank.

Legovini, Arianna, and Malte Lierl. 2009. *Impact evaluation in Fragile States*. BNPP Proposal. Washington, DC: World Bank.

Legovini, Arianna, Elizabeth Lule, and Leandre Bassole. 2008. *Africa HIV/AIDS Impact Evaluation Program*. Concept Note. Washington, DC: World Bank.

Legovini, Arianna, Muna Meky and Jee-Peng Tan. 2007. *Africa Program for Education Impact Evaluation*. Concept Note. Washington, DC: World Bank.

Mansuri, Ghazala, and Vijayendra Rao. 2010. *Policy Research Report on Community Driven Development and Local Governance*. Washington, DC: World Bank.

Martinez, Sebastian. 2007. *Health Results-Based Financing: SIEF Cluster Proposal*. Washington, DC: World Bank.

Mistiaen, Johan. 2010. *Assessing World Bank Investments in Microdata Generation: Preliminary Results*. Unpublished draft. Washington, DC: World Bank.

Neyman, J. 1952. *Lectures and Conferences on Mathematical Statistics and Probability*, 2nd edition. Washington, DC: US Department of Agriculture.

Nguyen, T.V. 2008. *Improving Management in Education: Evidence from a Randomized Experiment in Madagascar*. J-PAL Working Paper.
http://povertyactionlab.org/papers/93_Nguyen_Improving_Management.pdf.

Rao, Vijayendra. 2010. *Impact Evaluation*. Concept Note (unpublished). Washington, DC: World Bank.

Ravallion, Martin. 2008. *Evaluation in the Practice of Development*. Policy Research Working Paper. Washington, DC: World Bank.

Sakho, Cheikhou. 2009. *Note Conceptuelle D'évaluation D'impact de la Stratégie de Parrainage sur Conseil de Dépistage Volontaire*. Presentation at the HIV/AIDS Impact Evaluation Workshop, Cape Town, South Africa.

Shah, Shekhar, and Ariel Fiszbein. 2009. *Making Results Count: A Strategy for World Bank Monitoring and Evaluation Work on India*. Washington, DC: World Bank.

Torero, Maximo. 2006. *Estimating Impact of Rural Electrification Programs in Ethiopia*. Powerpoint presentation, IFPRI.
<http://siteresources.worldbank.org/EXTIMPEVA/Resources/Torero.ppt>.

World Bank. 2010a (forthcoming). *Améliorer la gestion de l'enseignement primaire à Madagascar - Résultats d'une expérimentation randomisée*. Africa Human Development Working Paper Series. Washington, DC: World Bank.

World Bank. 2010b (forthcoming). *Making Schools Accountable: What Works in the Developing World?*

World Bank. 2009a. *Annual Review of Development Effectiveness, 2009: Achieving Sustainable Development*. IEG. Washington, DC.

World Bank. 2009b. *Concept Note on the Development Impact Evaluation Initiative*. DIME. Washington, DC.

World Bank. 2006. *Impact Evaluation and the Project Cycle: Doing Impact Evaluation No. 1*. Washington, DC.

Annex A. DIME Governance Structure

DIME is a Bank-wide decentralized structure that counts with a governing body, a secretariat and, currently, collaboration from 22 separate units in the Networks and the Regions.

Steering Group

A DIME Steering Group (SG) was established in November 2008. The SG is composed of Director and Chief Economist level representatives from Networks, Knowledge, DEC, OPCS and IFC, and includes regional representatives on a rotating basis. The Chair is selected on a rotating basis with HDN, PREM and DEC serving for year 1, 2 and 3. The SG is responsible for:

- Providing the authorizing environment for DIME activities
- Prioritizing, overseeing and coordinating actions across networks
- Ensuring quality
- Defining and implementing strategic staffing
- Consolidating key partnerships and leading outreach activities
- Fund raising
- Annual progress reports to the Managing Director's office

DIME Secretariat

The DIME Secretariat works with networks and regions towards the adoption of effective mechanisms for the implementation of programs of impact evaluation that meet the objectives of the Bank-wide initiative. The Secretariat provides support and advisory services, including:

- Intellectual leadership and technical advice on the use of impact evaluation
- Small financial incentives to teams from RSB and TF resources
- Guidance to expand the thematic coverage of impact evaluation programs and adoption of good practices in technical support to project teams and quality assurance
- Pilot the development of country-based programs to networks and regions
- Capacity development and training, events and other dissemination activities
- Generation of IE products, meta-analysis, policy papers and corporate reports
- Liaison with international and local academic institutions
- Web and database information for the Bank impact evaluation activities
- Monitor and report on the Bank's impact evaluation activities

Networks

The Networks are responsible for managing thematic programs, to include:

- Coordination and technical support to project teams
- Thematic leadership and technical advice on the thematic priorities
- Integrate impact evaluation activities with other network business lines
- Training and other capacity development activities
- Thematic meta-analysis, policy papers and technical notes

- Events and other dissemination activities

Regions

The regions are responsible for implementation of the impact evaluation products, to include:

- Integrate impact evaluation in operations
- Policy dialogue
- Design and implementation of IE products
- Conduct training and other capacity development activities for internal and external resources

Research department

DEC provides technical resources to collect data and conduct analytical work. DEC is responsible for:

- Technical leadership and analytical services for thematic programs' activities and individual impact evaluation products
- Data quality and participation in data collection activities

Annex B. Description of Thematic Impact Evaluation Programs

I. Finance and Private Sector Program

While impact evaluation in financial and private sector has been rare, recent work demonstrates the potential that learning can have in informing the policy process and guiding the implementation of programs aimed at increasing private sector outcomes. The evaluations find, for example, that grants to entrepreneurs raise incomes and secure real returns to capital in Sri Lanka and Mexico (De Mel et al. 2008, McKenzie and Woodruff 2008). On business credit, individual lending outperforms group lending in the Philippines through faster client growth and stable default rates (Giné and Karlan, 2008), and welfare increases for loan recipients despite high interest rates (Karlan and Zinman 2008). Rainfall insurance fails to raise use of credit among Malawi farmers. (Giné and Yang 2009), and that product features matter for rainfall insurance take up (Cole et al. 2008). On business regulation, business registration reform increase registrations and employment in Mexico (Bruhn 2008, Kaplan et al. 2007) and bankruptcy reform in Colombia increases opportunities for firm reorganization and reduces liquidations (Giné and Love 2006).

Building on this experience, the Finance and Private Sector program aims to scale up evaluation in countries in Africa, Asia, Europe, and Latin America. The program focuses on understanding the constraints to growth and the mechanisms that are effective in releasing these constraints. Constraints include access to formality, managerial and technical skills, access to financial services, access to input and output markets, and protection under the law. Some of the questions the program will address are:

- Does formalization have positive returns, and what reforms and interventions increase the rate of small firm formalization? (Brazil)
- How does reduction in judiciary discretion improve firm access to protection under property and contract law? (Senegal)
- Do financial products change the growth prospects of firms? (Ethiopia, India, Turkey)
- Does trade facilitation help provide access to export markets? (South Africa) and are managerial skills needed to break export markets barriers? (Brazil)
- Is financial literacy conducive to intertemporal choices that are more conducive to welfare and economic growth? (Brazil, Uganda)
- Does market information improve market outcomes? (South Africa, Benin)

The program was launched in February 2010 in Dakar with the participation of Benin, Brazil, Cape Verde, Ethiopia, The Gambia, India, Malawi, Mauritius, Nigeria, Rwanda, Senegal, South Africa, Tanzania, Togo, and Uganda. The technical advisory team includes: Abhijit Banerjee (MIT), Alvaro Gonzalez (WB), Antoinette Schoar (MIT), Arianna Legovini (WB), Bilal Husnain Zia (WB), David McKenzie (WB), Dean Karlan (Yale), Elena Bardasi (WB), Leonardo Iacovone (WB), Markus Goldstein (WB), Miriam Bruhn (WB), Mattea Stein (WB), Sandra Sequeira (LSE), Shawn Cole (Harvard Business School), William F. Maloney (WB), and Xavier Giné (WB).

II. Human Development Programs

The programs in the human development network have been supported by a large grant by the Spanish and UK governments. The Spanish Impact Evaluation Fund is described in Box 9.

Box 9. The Spanish Impact Evaluation Fund (SIEF)

The Trust Fund for Impact Evaluation and Results-Based Management in Human Development Sectors—or Spanish Impact Evaluation Fund (SIEF)—aims to strengthen the Impact Evaluation of innovative programs to improve Human Development (HD) outcomes. Its ultimate goal is to enhance the effectiveness of development policies by strengthening the evidence base on the impact of programs affecting HD outcomes. With a \$16.3 million donation from the governments of Spain and the United Kingdom, the SIEF is the largest trust fund ever established in the World Bank focusing on impact evaluation.

Launched in July 2007, the SIEF program is managed by the Bank’s Human Development Network Office of the Chief Economist. SIEF resources support: (i) Prospective, rigorous impact evaluations of programs in 11 eligible HD and Sustainable Development sectors and 72 eligible developing countries across all regions; (ii) Intensive training programs for government counterparts, Bank staff, and staff of partner development agencies in IE methods; and (iii) Publication and dissemination of evaluation results through articles, meta-studies, and web-based materials.

The SIEF program supports the World Bank in building evidence from operations on “what works” to promote HD outcomes. Evaluations funded by the SIEF mainly focus on measuring valid estimates of the causal effects of development programs on HD outcomes. This is achieved through impact evaluations designed in such a way that outcomes for beneficiary groups can be compared with a valid control or comparison group.

At present, the SIEF is funding more than 50 impact evaluations, of which 36 are grouped in 7 “clusters”. These cluster studies aim to build communities of practice to generate evidence on how programs work across different country contexts. SIEF studies not grouped into clusters mainly focus on evaluating the impact of highly innovative interventions.

As of March 1, 2010, the SIEF has delivered 10 impact evaluation workshops in all developing regions in the world. More than 1,500 participants among evaluation experts, researchers and managers of programs affecting HD outcomes have been trained in the application and management of impact evaluations.

Improving the Accountability and Quality of Education

Efforts to improve education in both the developed and developing world typically focus on providing more inputs to schools—increasing spending along existing allocation patterns. But, substantial evidence shows that increased funding is not sufficient for improved learning outcomes. Incremental funds may be allocated to inputs that have weak impacts on student learning. In the United States, a tripling of real education spending per student since 1960 has been absorbed by higher teacher salaries and lower class sizes but has had no measurable impact on either student numbers or average student learning levels. Teachers and other education personnel (which typically represent 75 percent or more of education spending) may be poorly motivated to perform.

The 2004 World Development Report Making Services Work for Poor People argued that the underlying causes of such failures in basic service delivery in developing countries are weak accountability relationships between the state, service providers, and the citizens and

clients they serve. In the education sector, efforts in both developed and developing countries to strengthen these accountability relationships through system reforms have been numerous.

This program of impact evaluation answers the following questions:

- What type of school-based management improves service delivery and learning outcomes?
- Does providing parents with better information about school quality have an impact on school performance?
- Does the type of teacher contractual arrangements matter for schooling access and student learning?

Early Childhood Development

Leveling the playing field must start in early childhood. Early nutrition and mental stimulation, along with maternal health, have significant impacts on an individual's long term physical, cognitive, and emotional development. Without proper nutrition and necessary care in the first five years of life, a child's educational, and future earning potential can be severely disadvantaged.

Under-investment in early childhood development results in poverty, poor health and nutrition, and deficient care; an estimated 200 million children under the age of five do not reach their cognitive development potential as a result of these factors. Under-nutrition, iron deficiency, iodine deficiency and other nutritional factors stunt the cognitive abilities of children. Stunting at infancy is correlated with lower educational achievement. Likewise, health risks posed by infectious disease and environmental factors negatively affect child development. Remedying the causes of such insufficient child development at an early age is critical to improving the potential of children.

Finding viable and low cost options to implement early childhood interventions is critical for countries' ability to scale up early childhood interventions and secure the long-term productivity of their labor force. The impact evaluation program focuses on identifying the packages and delivery mechanisms to make ECD a viable option; questions include:

- What is the best combination of cognitive stimulation, early nutrition and health interventions for improving child nutrition, anthropometric measures, and cognitive achievement? And how much does this matter for school readiness?
- What is the best way to ensure these packages are delivered effectively?
- What program alternatives, such as parental enrichment, or community and center-based early childhood intervention are most effective? And what are the complementarities across these different models?

Making Health Systems Work

Lack of services - insufficient geographical coverage, inadequate numbers of health staff, lack of motivation or training among healthcare workers, and lack of health and nutritional supplies – is a major factor in poor health outcomes. Constraints on public budgets and human resources for health and education mean that governments need cost-effective

ways of drawing on private and non-profit sectors for delivery of services and getting the best performance out of publicly paid providers.

The World Bank's World Development Report (WDR) 2004 argued incentives for delivering high quality care and responsiveness to patients need to be strengthened. Pay-for-performance contracts offer a means by which the government can align the payment structure of health centers with service outputs and health outcomes. These contracts are intended to motivate better patient outcomes by tying providers' remuneration to specific targets for service delivery quality and quantity. The program will focus on some of the very important questions that will help countries save the lives of mothers and children:

- How does the introduction of case-based payment for patient services impact the quality and performance of county hospitals?
- What pay-for-performance schemes increase the quantity and quality of services provided? And what improvements in health results can be expected?

HIV/AIDS

The global AIDS epidemic is fueled by risky sexual behavior. Over 80 percent of HIV infections occur through sexual contact with an infected partner and could have been avoided through the adoption of safe sexual behavior such as condom use, reduction in concurrent partnerships, abstinence and type of sexual interaction. Prevention programs appear to have been fairly successful in increasing awareness and knowledge but evidence on the link to changes in sexual behavior is weak. The question of whether HIV prevention is effective and whether the right strategies are being implemented is on everyone's mind. Are the economic and behavioral factors that drive risky sexual behavior being addressed? This is the focus of impact evaluation in HIV prevention:

- What interventions substantially reduce risky sexual behavior?

For example, experimental evidence that male circumcision reduces the risk of infection by 60 percent has been instrumental in getting national HIV agencies to start developing circumcision programs.

On the treatment side, the emphasis had been on making Antiretroviral Therapy (ART) available and increasing the number of HIV infected individuals on treatment. However, ART is only beneficial when patients have very high levels of adherence to the treatment. Without high adherence patients could be harmed by the medication and millions of dollars wasted. Learning how to secure high adherence to treatment is an enduring challenge that requires testing of multiple competing strategies and answer the question of:

- What interventions are successful in ensuring patients' high level of adherence to ART?

Understanding the answer to this question is critical to save lives, increase productivity and ensure the effectiveness of large public expenditures in ART.

Malaria

Together with the World Bank's Booster Program for Malaria Control, the Malaria Impact Evaluation Program (MIEP) focuses on critical knowledge gaps on the effectiveness of alternative modes of delivery, and channels of communication.

On prevention and behavioral change, countries are engaged in major distribution campaigns of Long-Lasting Insecticide-Treated Nets (LLINs) yet the gap between ownership and usage is large. D.R. Congo, India, and Nigeria are testing different communication strategies to find out what it takes to increase wide use of LLIN and eliminate malaria from entire communities.

Improved and accessible treatment to Artemisinin combination therapy (ACTs) is a challenge. The drug is very effective in treating malaria, but it is also very expensive. How can we ensure that ACT is available to those that need it at a price they can afford? India, Nigeria and Zambia are investigating the role of subsidies, public-private partnerships, community engagement, and supply chain interventions in addressing bottlenecks.

Yet, protecting against ACT misuse and drug resistance goes beyond the availability and affordability of test kits. It requires a change in the behavior and attitudes of medical staff, pharmacies and patients. Zambia and India are investigating the role of subsidized rapid diagnostic tests and public-private partnerships in ensuring the responsible use of ACTs.

The health systems also face significant constraints. Nigeria, D.R. Congo and Zambia are testing providers' incentives to improve service performance.

More innovative ideas are on the horizon. Kenya and Senegal are investigating the role of preventive treatment (IPT) in schools. Eritrea, having eliminated malaria mortality through a full-fledged program of malaria control, is now testing the use of indoor residual spraying to take one more step toward eradication.

Active labor market programs

Particularly in the aftermath of the financial crisis, joblessness and underemployment are one of the most challenging economic and social problems policymakers face in developing countries. Policymakers throughout the world struggle to find effective programs that increase employability and earnings of individuals as well as the quality of the jobs offered. As disincentives and dependencies of social transfers (including unemployment insurance) have become better understood, active labor market programs/youth employment programs have raised attention as an attractive, and often complementary, option.

ALM programs include a wide range of activities, intended to foster the quality of labor supply (e.g., training), to increase labor demand (e.g., public works); or to improve the matching of workers and jobs (e.g., job search assistance, employment agencies). They usually target a specific set of disadvantaged individuals that range from the unemployed and unskilled adults, to women or first time job seekers. However, the effectiveness – and efficiency – of ALMP/YE programs in reaching their target groups, remains poorly documented. Despite the urgent policy demand, surprisingly few interventions have been accurately evaluated in the context of developing countries and good practices are still to be defined.

The impact evaluation program seeks to improve our knowledge on three core policy questions. First, do ALMPs effectively improve the employability and productivity of the targeted groups in the population? Second, which interventions are most cost effective and for whom? Third, how can these programs be better designed (or implemented) to effectively reach those who need them the most?

The program facilitates support to World Bank teams to enhance the design and implementation of impact evaluations of ALMPs. It promotes the use of harmonized indicators, produces knowledge products and organizes dissemination events to build a community of practice on ALMPs. The cluster is currently providing financial support through the Spanish Impact Evaluation Fund on eight impact evaluations (Box).

Conditional Cash Transfers

Conditional Cash Transfers (CCT) have become an increasingly popular strategy for poverty reduction programs. Though used most extensively in Latin America, CCT programs are now being implemented across the globe. The idea behind CCT programs is simple – cash transfers are provided to households that meet certain specific conditions. In other words, for a household to receive a cash transfer they must undertake certain activities (like regular health exams for children or ensuring that school-aged children go to school). The objective is to make short-term income transfers contribute to the longer-term objective – protecting the health and schooling of low-income children as a strategy for more sustained poverty reduction.

Mexico's Health, Nutrition and Education (PROGRESA) program was launched in 1997 to combat the country's high poverty rate and to replace a food subsidy and other poverty programs that were considered ineffective. As the program was phased in with a strategy of randomly assigning eligible communities into the program over a three year period, by 1999 it was possible to rigorously evaluate the impact of the program. The evaluation found that the program produced noteworthy increases in school enrollments, especially in middle school enrollment, declines in levels of child malnutrition and illness, and reductions in poverty. Since then, the program has been expanded, weathered multiple political changes and now serves 20 million people - one-fifth of Mexico's population.

Despite the accumulated experience with CCT programs, there remain critical questions regarding CCT program design, implementation and context. The goal of this program is to answer these critical questions.

- Do CCT programs need conditionality in order to improve outcomes? If so, what types of conditions work best?
- What are the most cost-effective levels of transfers and to what populations?
- Does it matter who in the family receives the conditional cash transfer?
- How much does the quality of supply (availability or quality of local schools or health clinics) affect the outcomes from demand-side incentives like CCTs?

III. Sustainable Development Programs

Agricultural Adaptations

75 percent of the world's poor depend on rain-fed agriculture for their livelihoods. Climate variability is set to increase farmers' vulnerability in most of the developing world. In light of the twin challenges of food insecurity and climate change,

- What combination of prices and information will elicit the rapid technology adoption needed to secure high returns to irrigation and other rural infrastructure and resilience to climate change?
- What structure of incentives will induce a sustainable use of land and water resources?
- What gender-specific interventions will improve economic returns to agriculture and agricultural infrastructure?

While technologies are available that can provide farmers with effective coping strategies, creating incentives for farmers to adopt new, often expensive production inputs presents a substantial operational challenge. Impact evaluation provides the analytical tools to address these operational issues. By testing side-by-side alternative implementation arrangements designed to enhance farmers' incentive to adopt new technology, impact evaluation will provide scientific evidence on what interventions deliver results on the ground. AADAPT will contribute to transforming the rural development sector strategy into effective and workable solutions on the ground. AADAPT will support rigorous learning along four pillars: 1. Land and water management; 2. Improving access to markets and supportive rural infrastructure; 3. Food security and vulnerability; and 4. Agricultural technology. The program has a strong gender focus that permeates all evaluations. The gender team collaborates with DIME to support the design of rigorous gender aware impact evaluations and measure gender-disaggregated impacts of specific operations.

Access to infrastructure

Urban Upgrading

An estimated one-third of all urban residents live in informal settlements or slums—the vast majority in developing countries. Globally, almost one billion people live in slums according to United Nations estimates. Conditions in such areas vary widely from dismal, temporary shelter in squatter settlements to relatively well-constructed, informal housing that may persist for many decades. Common characteristics include uncertain tenure status, poor basic services such as water and sanitation, low-grade construction and overcrowded living conditions. Apart from physical deprivation, slum dwellers often face more subtle disadvantages such as poor labor market integration and the social stigma attached to an inferior residential location, environmental hazards, crime and violence, and a breakdown of traditional family and community safety nets typically found in rural villages. Children living in slums are deprived of access to good quality education and health services, which are not located in reasonable proximity of these settlements.

Improving the lives of slum dwellers is a high priority for national and city governments and the international community. The Millennium Development Goals, for instance,

advocate significant improvements in the lives of at least 100 million slum dwellers by 2020. Over the last several decades, strategies of national governments and development agencies to achieve better living conditions of slum dwellers have included a range of urban upgrading activities such as sites and services (including both infrastructure interventions at the community and household level, as well as social interventions such as job training, day care and community development), resettlement to new housing developments, housing subsidies, and land titling. Most programs tend to be multi-dimensional including several activities.

While there is an urgent need to scale up interventions that improve the quality of life for slum dwellers, there is little clarity on (a) the types and composition of interventions that are most effective; (b) the sustainability of alternate programs (c) their relative cost effectiveness (d) fiscal impact and (e) the citywide consequences of these interventions. The program focuses on questions such as,

- How does in situ upgrading compare to housing development programs?
- What is the impact of housing improvements on health and labor outcomes?
- What is the value added of different types of infrastructure upgrade?
- Is titling a necessary component for upgrading projects to work?
- What private benefits can be expected for slum dwellers?
- What social community wide impacts do upgrading programs have, such as improvements in perceptions of social equity, social capital, and levels of crime and violence?
- Do housing vouchers improve labor market integration?
- Are housing subsidies effective in increasing home investment?

Rural Infrastructure

Roads. Poor quality roads are frequently perceived as an impediment to economic development and poverty reduction. Lack of maintenance, erosion and flooding can disrupt or erase the utility of existing roads. An inferior roads network increases transport costs that, in turn, constrain investment and economic activity. Approximately 900 million people in rural areas live without access to all-season-roads. This gap constitutes a serious hindrance to economic and social opportunities of those living in these areas. Governments in developing countries devote significant efforts to expand and improve rural roads, often with international funding. The International Development Association (IDA) alone allocates approximately \$1 billion a year to rural roads projects.

Impact evaluations of road infrastructure are complex because of the economy-wide effects that roads create. Roads influence a wide array of economic and social activities. Acting through lowered transport costs, roads might promote market activities, the availability and use of social services; affect the division of labor inside and outside the household; etc. A thorough evaluation of all these effects is necessary in order to assess the contribution of this type of investment on the welfare of the population. The program's questions are:

- What is the mechanism by which roads improvements can be translated into improved living conditions and lowered inequality?
- How do local conditions affect impacts of road investments?

- What investments (i.e. complimentary infrastructure, productive activities) and services (i.e. transport options) interact with roads to improve development outcomes?
- What management schemes are most effective for maintaining good-quality roads?

Electrification. More than 1.6 billion people in developing countries lack electricity. Nearly all of these people rely on traditional biomass energy sources to meet their heating and cooking needs. Electrification can improve conditions for human capital development through lighting, as potential studying and reading time increase. By promoting a more efficient use of inputs, electricity also has the potential to increase the productivity of household enterprises. Moreover, electricity is expected to improve quality-of-living standards in communities through the introduction of street lighting, enhanced access to information sources, expanded health services due to refrigeration, and improved learning conditions in schools.

However, little has been done to measure the causal impact of electrification on living standards, economic activity, and human capital investment. The interventions included in this cluster include grid extension, micro-grid technologies, and renewable energy. As electricity is delivered to rural communities for the first time, the studies in this theme will measure the impact of access to electricity. In addition, alternative off-grid technologies will be piloted in diverse settings in order to establish comparable estimates of multiple methods of delivering electricity to rural communities. To capture the effect of price on usage, some studies will estimate willingness to pay through experimental means, whereby the price that a household pays for electricity is varied at random.

Adaptations in Water Resource Management

Program in development.

Mitigation in Energy

Program in development.

Adaptations in Forestry

Program in development.

IV. Poverty Reduction and Economic Policy Programs

Gender¹⁶

The DIME-GAP Collaboration is a Model to Mainstream Gender in World Bank Operations. The Gender Action Plan (GAP) – Gender Equality as Smart Economics – seeks to advance women economic empowerment in World Bank client countries through intensification of gender mainstreaming in the economic sectors. The GAP focuses its actions at the Policy Level to make markets (labor, land, product and financial) work for women, and at the Agency Level to empower women to compete in those markets. GAP’s guiding principles include selective coverage, evidence base, results orientation, and reliance on incentives rather than mandates. In this context, GAP has issued several calls for proposals for engendering operations, research and impact evaluations. In addition to that, there have been efforts in capacity building and investments in gender statistics. Agriculture is one of

¹⁶ This section was provided by Mayra Buvinic and Rui Benfica (PRMGE).

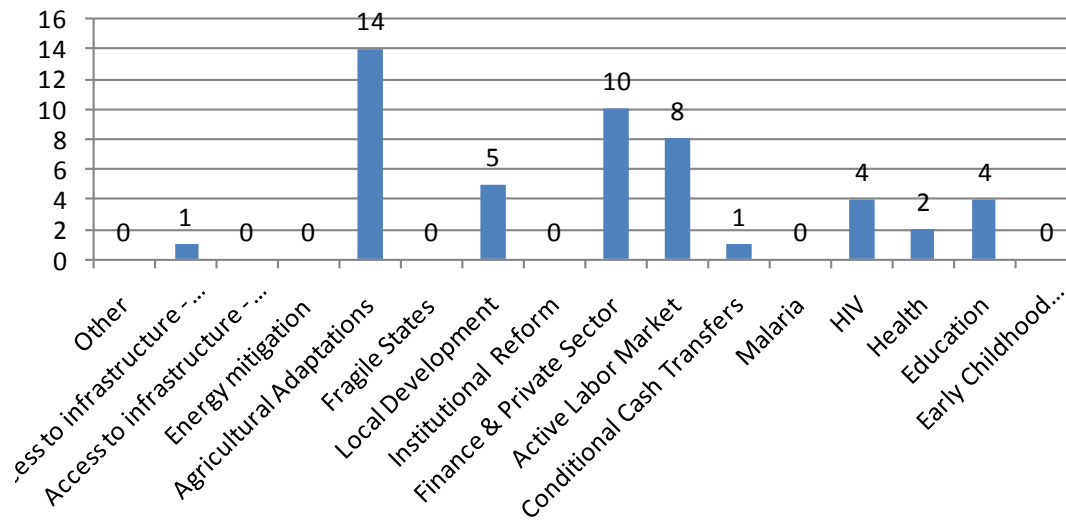
GAP's sectors, a sector where women have an overwhelming participation but face acute constraints and unequal access to resources, rights and opportunities, that cause inefficiencies in household production, slower economic growth and poverty reduction.

The GAP and the Development Impact Evaluation (DIME) are partnering in the context of the Agricultural Adaptations – AADAPT, and the Finance and Private Sector—DIME-FPD, DIME's programs of impact evaluations that support country programs getting answers, in real time, to their most pressing operational questions. The model provides a good platform to rigorously identify gender influences and effects in Bank operations and to identify relevant gender concerns and integrate them into project design, management, monitoring and evaluation. GAP partners with DIME through funding to (1) Cross-country Impact Evaluation Workshops with Task Team Leaders and Project Teams where gender is integrated in learning and concept note design; and (2) Incentive funding to support coordination and ensure gender integration in project implementation. The GAP has contributed 1.7 million to the effort in the Africa Region, South Asia and Latin America regions.

There are three key elements in the AADAPT model that justify the partnership with DIME. First, it is an appealing “customized capacity building” model that, unlike traditional capacity building, brings together learning, research and operations from design to implementation and evaluation of impacts. Second, it is client centered, inclusive and operationally relevant. In fact, it brings together Bank Staff and counterparts throughout the process, links learning to design of operationally relevant gender aware evaluation of projects through experimenting with gender differentiated interventions and measuring gender disaggregated impacts of interventions, and is capable of influencing operations through feedback from research/Impact Evaluation findings, therefore helping governments retool their interventions in specific contexts to bridge the gender gap. Finally, it is an effective way to mainstream gender in line with GAP objectives and philosophy as it uses evidence base to influence the integration of gender in Bank operations, relies on incentive funding to strengthen gender mainstreaming in Bank operations, and generates substantial gender disaggregated data to support broader policy analysis in client countries.

The GAP-DIME collaboration model in AADAPT and DIME-FPD has proven to be effective in increasing the likelihood of integration of gender in Bank operations and impact evaluations in ARD and FPD. Using a model that links training to operations and highlights and makes the business rationale for gender equality is a very effective strategy. While incentive funding plays a key role, it is not enough to motivate teams. Continued advice and follow up with teams is fundamental to secure the integration of gender in design and sustain it through implementation and analysis. This model can be successfully replicated in other sectors in all regions to place gender at the center of the research and operational agenda.

En-gendered impact evaluations



Institutional Reforms

Since the late 1990's, the Bank has increasingly recognized the critical importance of well performing public institutions and good governance for development and poverty reduction. Despite the increasing widespread efforts to promote such reforms, the impacts of these interventions remain largely unexplored.

Moving from a normative dialogue to one based on practical recommendations is of critical importance for advancing the governance agenda at greater speed. Building the body of evidence that will help governments improve the success of Governance and Institutional reforms is the focus of this program of impact evaluation.

Governance and institutional reforms may include initiatives to: (i) deregulate the economy, (ii) decentralize government structure to bring services closer to the people, (iii) strengthen community action and public accountability, (iv) strengthen "accountability" institutions such as the judiciary, audit bodies, and anti-corruption commissions, and (v) reform public sector management in areas such as customs, tax administration, or the civil service.

Despite the diversity in the nature of the different interventions, a general question pertaining most of these interventions is how these types of reforms ultimately affect service delivery, which is the most direct channel through which Governance affects poverty.

Some more specific questions this program is currently trying to answer are as follows:

- Do programs aiming to increase the access to justice in fragile environments increase beneficiaries' ability to claim rights, and resolve disputes through non-violent means? Do they improve household welfare? Do they increase the reporting of corruption and bribery?

- Are Access to Information Reforms effective at increasing accountability and transparency of the Government and do they have an impact on population's perception of Government?
- In a fiscally decentralized environment, do reforms in the electoral system (e.g. direct elections of district government heads) improve government accountability and the delivery of public services?
- Can satisfaction survey data serve as a reliable indicator of the quality of health and education services?

V. Multi Sector Programs

Local Governance

As decentralization of government decisions and fiscal resources has become a powerful mechanism for reform, questions like these are critical to national development strategies. All the more important are reliable, precise answers. What works in one country may fail in another. Hence, the best strategies build on continuous learning, testing of innovative approaches and gradual improvement of programs, to find out what works best in each context.

Learning about the effectiveness of local governance support is a difficult challenge. Many of the intended outcomes require substantial effort and methodological sophistication to be measured in a sensible way. This is the case for abstract concepts such as 'social cohesion', 'empowerment', or 'accountability', and even for concrete outcomes, such as the implementation performance of local investment projects.

Impact evaluations play a key role in producing dependable evidence and comprehensive data on program effectiveness. This program helps governments access the best expertise for evidence-based improvement of local governance programs and answer questions such as:

- Does local participation establish a more accountable and effective local governance? Does it improve the legitimacy of local governments? Does it enhance the maintenance and sustainability of public services? Does it increase access to resources and information?
- What type of information can effectively enhance local accountability?
- Does competition for funding enhance the quality of local decisions?
- How much mobilization and facilitation is needed to make participation productive?

Fragile states

The World Bank's fragile states strategy prioritizes community-driven development and local governance operations as a response to situations of deteriorating governance and political instability. These programs hold the promise of delivering services to the poor and improving the relationship between citizens and public institutions, without necessarily touching politically sensitive questions of governance at the national level.

The objective of this program is to investigate the impact of local development programs on the relations between citizens and authorities in situations of state failure or deteriorating governance. The impact evaluation studies in this program will focus on

questions of legitimacy and utilization of public institutions, public service delivery, local collective action, social capital and trust in elected officials. They will answer questions such as:

- Can local development efforts improve the relations between citizens and authorities in situations of state failure or deteriorating and build social capital and trust in elected officials?
- How can centrally and community-driven interventions be complementary in the reconstruction process?
- How can vulnerable groups partake in the state and peace building process?
- What political and social accountability mechanisms are most effective in a fragile state?

All impact evaluations will be carried out in partnership with governments and implementing agencies, strengthening their capacity for evidence-based policy decision-making and critical reflection of policies. They will produce reliable evidence on the impact of local development programs, enabling fragile states governments to identify realistic measures to improve local stability, and to articulate those vis-à-vis the donor community. The results are expected to inform specific project implementation decisions, country strategies, and contribute to upstream work on Bank responses to deteriorating political situations and de-facto governments.

Annex C. Academic partners

DIME works in a collaborative mode with a large number of academic institutions. Many of the researchers contribute part of their time pro-bono. Many others are hired as short term consultants to work on specific products. All contribute their expertise and efforts to take development to the next level. Here is a list of institutions these researchers are associated with.

Aarhus School of Business (Denmark), Abdul Latif Jameel Poverty Action Lab (USA), AEI-Brookings (USA), Bangladesh Institute of Development Studies (Bangladesh), BASICS II (Honduras), BASICS/Honduras (Honduras), Bocconi University (Italy), Brown University (US), Cellule de Lutte contre la Malnutrition Dakar (Senegal), Center for Global Development (US), Center for Health Statistics and Information (USA), CERDI-CNRS (France), Chicago GSB (USA), CNLS (Rwanda), Columbia University (USA), Cornell University (USA), Dartmouth University (USA), Deutsches Institut für Wirtschaftsforschung (Germany), Duke University (USA), Econometría Consultores-Bogotá (Columbia), Emory University (USA), ESA Consultores (Honduras), FAID (Germany), Florida State University (USA), Fundação João Pinheiro (Brazil), George Washington University (USA), Grupo de Análisis para el Desarrollo (Peru), Harvard Business School (USA), Harvard University (USA), IFPRI (International Organization), IGIER (Italy), Innovations for Poverty Action (USA), Institute for Fiscal Studies (UK), Institute for Empirical Research in Political Economy (Benin), Institute for Research on Evaluating Public Policy (Italy), Institute of Social Studies (The Netherlands), Instituto APOYO (Peru), Instituto Colombiano de Bienestar Familiar (Colombia), Inter-American Development Bank (International Organization), Iowa State University (USA), IPEA (Brazil), IPS (USA), ISID (USA), JHU School of Public Health (USA), Johns Hopkins University (USA), Kennedy School of Government (USA), Khon Kaen University (Thailand), London School of Economics (UK), London School of Hygiene and Tropical Medicine (UK), Massachusetts Institute of Technology (MIT) (USA), National Bureau of Economic Research (USA), New York University (USA), Office of Population Studies-San Carlos University (Philippines), Pomona College (USA), Peradeniya University (Sri Lanka), Philippine Institute for Development Studies (Philippines), Population Council (USA), Pratham (India), Princeton University (USA), RAND Corporation (USA), Stanford University (USA), Stockholm University (Sweden), The Brookings Institution (USA), The Geneva Graduate Institute (Switzerland), Unibanco (Brazil), UNICEF (International Organization), Università di Padova (Italy), Universidad de San Andrés (Argentina), Universidad Iberoamericana (Mexico), Universidad Torcuato Di Tella (Argentina), Universidade Federal de Pernambuco (Brazil), Universitat Pompeu Fabra (Spain), Université d'Abomey-Calavi (Benin), Université du Québec à Montréal (Canada), University College London (UK), University of California, Berkeley (USA), University of California, Davis (USA), University of California, Los Angeles (USA), University of California, Riverside (USA), University of California, San Diego (USA), University of Gadjah Madah (Indonesia), University of Illinois (USA), University of Michigan (USA), University of Minnesota (USA), University of North Carolina (USA), University of Oklahoma (USA), University of Pennsylvania (USA), University of Sao Paulo (Brazil), University of St. Louis (USA), University of Tokyo (Japan), University of Toulouse (France), University of Washington (USA), University of York (UK), UWI-Jamaica (Jamaica), Vanderbilt University (USA), W.E. Upjohn Institute for Employment Research (USA), Wageningen University and Research Centre (Netherlands), Waikato University (New Zealand), Washington University (USA), Wellesley College (USA), World Health Organization (International Organization), and Yale University (USA).

Annex D. DIME Data Catalogue¹⁷

DIME has established collaboration with the DEC Data Group to establish a DIME data Catalogue. Why establish a DIME data catalogue?

Microdata files provide the information required to conduct impact evaluations. Microdata are information about individual respondent units that are collected via sample surveys, censuses, administrative records and various management information systems.

Respondent units include persons, households, firms and service delivery facilities such as schools and health centers. The World Bank provides substantial assistance, both technical and financial, to support the collection and analysis of microdata in developing countries. Three key objectives of documenting and archiving microdata files used for impact evaluations are to:

(a) **Improve Returns on Impact Evaluation Investments:** Microdata constitute irreplaceable assets that are costly to acquire and which must be managed in a way that encourages their widest possible use and re-use. Data have value beyond the purpose for which they were originally collected (“repurposing” of data). Many surveys are conducted but remain hidden to potential users. By creating a central repository and maintaining on-line searchable catalogs of microdata and metadata—data about microdata—the DIME data archive will contribute to make existing data more visible. Dissemination of impact evaluation metadata and microdata (when possible) will promote learning and facilitate improvements in impact evaluation practices and techniques.

(b) **Enhance Transparency and Accountability:** Good information on how microdata and impact evaluation results are produced facilitates understanding, replication and quality. It is important to document the microdata files on which the impact evaluation results are based as well as the process and analytical methods used to generate them. Good practice for documenting impact evaluations is to comply with the replication standard which, succinctly, can be stated as follows: sufficient micro- and metadata exists with which to understand, evaluate, and build upon impact evaluation analysis if a third party could replicate the results without any additional information. Without proper descriptions of the design of the survey and the methods used when collecting and processing the data, the risk is high that the users might misunderstand or even misuse them. Microdata files in the DIME data archive will be clearly and fully documented in accordance with international standards, the data files will contain no surprises, and authorized users will be able to work the dataset with relatively little start-up time.

(c) **Safeguard the Knowledge Base:** Micro-datasets can be damaged or lost because of human error, because of technical problems that lead, for example, to the corruption of data files, or because of disasters such as fire or flood. New technologies can also render old data unreadable, because of either hardware or software advances. The DIME data archive will develop a data management plan that will include standard procedures for ensuring

¹⁷ This Annex was provided by Olivier Dupriez and Johan Mistiaen of the DEC Data Group.

the physical security of its resources together with associated backup arrangements for minimizing the impact of adverse events.

What will the DIME data catalogue contain? There are two main types of material that constitute ideal documentation for a DIME dataset: explanatory and contextual material.

Explanatory Material

This represents the minimum of material to create and preserve, and can be described as the material required to ensure the long-term viability and functionality of a dataset. Full understanding of the dataset and its contents cannot be achieved without this material.

Information about the data collection methods

This information describes the data collection process, whether it is a survey, the collection of administrative information, or the transcription of a document source. It should describe the instruments used, the methods employed, and how these were developed. If applicable, details of the sampling design and sampling frames should be included. It is also extremely useful to include information on any monitoring process undertaken during the data collection as well as details of quality controls.

Information about the structure of the dataset

Key to this type of information is a detailed document describing the structure of the dataset and including information about relationships between individual files or records within the study. It should include, for example, key variables required for unique identification of subjects across files. It should also include the number of cases and variables in each file and the number of files in the dataset. For relational models, a diagram showing the structure and the relations between the records and elements of the dataset should be constructed.

Technical information

This information relates to the technical framework and should include: the computer system used to generate the files; the software packages with which the files were created; the medium on which the data was stored, and a complete list of all data files present in the dataset.

Variables and values, coding and classification schemes

The documentation should contain a full list describing all variables (or fields) in the dataset, including a complete explanation and full details about the coding and classifications used for the information allocated to those fields. It is especially important to have blank and missing fields explained and accounted for. It is helpful to identify variables to which standard coding classifications apply, and to record the version of the classification scheme used - preferably with a bibliographic reference to that code.

Information about derived variables

Many data producers derive new variables from original data. This may be as simple as grouping raw age data (age in years) according to groups of years appropriate for the needs of the survey, or it may be much more complex and require the use of sophisticated algorithms. When grouped or derived variables are created, it is important that the logic for the grouping or derivation be clear. Simple grouping, such as for age, can be included

within the data dictionary. More complex derivations require other means of recording the information. The best method of describing these is by using flow charts or accurate Boolean statements. It is recommended that sufficient supporting information be provided to allow an easy link between the core variables used and the resultant variables. We would also recommend that the computer algorithms used to create the derivations be saved together with information on the software.

Weighting and grossing

Weighting and grossing variables need to be fully documented, explaining the construction of the variables with a clear indication of the circumstances in which they should be used. The latter is particularly important when different weights need to be applied for different purposes.

Data source

Details about the source the data is derived from should be included in some details. For example, when the data source is made up of responses to survey questionnaires, each question should be carefully recorded in the documentation. Ideally, the text will include a reference to the generated variable(s). It is also useful to explain the conditions under which a question would be asked to a respondent, including if possible, the cases to which it applies, and ideally, a summary of response statistics.

Confidentiality and anonymization

It is important to note if the data contains any confidential information on individuals, households, organizations or institutions. Whenever this occurs, it is recommended to record such information together with any agreement on how to use the data - for example, with survey respondents. Issues of confidentiality may restrict the analyses to be undertaken or the results to be published, particularly if the data is to be made available for secondary use. If the data was anonymized to prevent subjects' identification, it is wise to record the anonymization procedure and its impact on the data. Such modification may restrict subsequent analysis and an indication of it is useful.

Contextual information

This provides users with material about the context in which the data was collected, and how it was put to use. This type of information adds richness and depth to the documentation, and enables the secondary user to fully understand the background and processes behind the data collection exercise. This also forms a vital historical record for future researchers.

Description of the originating project

Details should be provided about the history of the project, or about the process that gave rise to the dataset. This should offer information on the intellectual and substantive framework. For example, the description could cover topics such as: why the data collection was felt necessary; the aims and objectives of the impact evaluation; who or what was being studied; the geographic and temporal coverage; publications or policy developments it contributed to or that arose as a response, and any other relevant information.

Provenance of the dataset

This information relates to aspects such as the history of the data collection process, changes and developments that occurred in the data themselves and the methodology, or any adjustments made. The following can be provided as well: details of data errors; problems encountered in the process of data collection, data entry, data checking and cleaning; conversion to a different software or operating system; bibliographic references to reports or publications that stem from the study, and any other useful information on the life-cycle of the dataset.

Serial and time-series datasets, new editions

For repeated cross-section, panel or time-series datasets, it is extremely helpful to obtain additional information describing, for example, changes in the question text, variable labeling or sampling procedures.

