

Corporate Evaluation

IDB's Impact Evaluations: Production, Use, and Influence

IDB's Impact Evaluations: Production, Use, and Influence

Office of Evaluation and Oversight (OVE)

October 2017





This work is distributed under a Creative Commons license <https://creativecommons.org/licenses/by-nc-nd/3.0/us> (CC BY-NC-ND 3.0 US). You are free to share, copy and redistribute the material in any medium or format, Under the following terms:



Attribution - You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.



Non-Commercial - You may not use the material for commercial purposes.



No Derivatives - If you remix, transform, or build upon the material, you may not distribute the modified material.

No additional restrictions - You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

The link provided above includes additional terms and conditions of the license.

© **Inter-American Development Bank, 2017**

Office of Evaluation and Oversight
1350 New York Avenue, N.W.
Washington, D.C. 20577
www.iadb.org/evaluation

RE-512-1
October 2017

ACRONYMS AND ABBREVIATIONS

ACKNOWLEDGEMENTS

EXECUTIVE SUMMARY

1. INTRODUCTION	1
2. THE PRODUCTION OF IMPACT EVALUATIONS IN IDB.....	9
A. Analysis of Production	9
1. How many IEs have been proposed and what is their status?	9
2. How have the processes for origination, design and implementation worked?	15
3. How much do IEs cost?	20
4. What are good practices in financing impact evaluations and knowledge generation in general?	24
B. What is the quality of completed evaluations and the designs of ongoing ones?	25
3. USE AND INFLUENCE OF IDB'S EVALUATIONS	31
A. Use and Influence of Impact Evaluations as Knowledge	31
1. To what extent are IEs used internally in IDB?	31
2. How are IDB IEs used by clients?	33
3. What is the broader reach of IDB IEs?.....	36
B. Can IE help in delivering projects?	37
4. CONCLUSIONS AND RECOMMENDATIONS.....	41

NOTES

REFERENCES AND BIBLIOGRAPHY

Available at www.iadb.org/evaluation

ANNEX I	Construction of the IDB Impact Evaluation Database
ANNEX II	Basic Statistics of Impact Evaluations at the IDB
ANNEX III	Impact Evaluation in Other Organizations – DIME, JPAL, and MCC
ANNEX IV	Best Practices in Financing Knowledge
ANNEX V	CLEAR Quality Assessment of IDB Impact Evaluations
ANNEX VI	Use of Evidence in Policymaking in Latin America and the Caribbean: The case of Impact Evaluation at the IDB
ANNEX VII	Do Impact Evaluations Help Deliver IDB Projects?
ANNEX VIII	Analysis of the Use of Impact Evaluation in the Sector Framework Documents

ACRONYMS AND ABBREVIATIONS

ADB	Asian Development Bank
AfDB	African Development Bank
CCB	Country Department Caribbean Group
CCT	Conditional cash transfer program
CID	Country Department Central America, Mexico, Panama and Dominican Republic
CSD	Climate Change and Sustainable Development Sector
DEF	Development Effectiveness Framework
DEM	Development Effectiveness Matrix
DIME	Development Impact Evaluation Initiative
EDU	Education Division
ESW	Economic and sector work
GCI-9	Ninth General Capital Increase
GDI	Gender and Diversity Division
IDB	Inter-American Development Bank
IE	Impact evaluation
IFD	Institutions for Development Sector
IMF	International Monetary Fund
INE	Infrastructure and Energy Sector
INT	Integration and Trade Sector
J-PAL	Abdul Latif Jameel Poverty Action Lab
KNL	Knowledge & Learning Sector
LAC	Latin America and the Caribbean
M&E	Monitoring and evaluation
MCC	Millennium Challenge Corporation
MIF	Multilateral Investment Fund
MDB	Multilateral development bank
OVE	Office of Evaluation and Oversight
RCT	Randomized controlled trial
RES	Department of Research and Chief Economist
RND	Environmental, Rural Development and Disaster Risk Management Division
SCL	Social Sector
SFD	Sector Framework Document
SPD	Office of Strategic Planning and Development Effectiveness
SPH	Social Protection and Health
TC	Technical cooperation
VPC	Vice Presidency of Countries
VPS	Vice Presidency of Sectors

This document was prepared under the guidance of Cheryl Gray (OVE Director) by a team composed of Anna Crespo and Oliver Azuara (team leaders), Cesar Bouillon, Veronica Gonzalez Diez, Alejandro Palomino, Florencia Alejandre, Johanan Rivera, and Maya Jansson.

OVE is grateful to IDB staff for their constructive contributions to the evaluation process, and to the individuals from other institutions and executing agencies who were surveyed and interviewed in the context of this study.



The introduction of the Development Effectiveness Matrix, has strongly incentivized the production and use of Impact Evaluations.

© IDB

Executive Summary

During the last decade the Inter-American Development Bank (IDB, or the Bank) has invested heavily in promoting impact evaluation (IE) as a tool to assess its effectiveness. In favoring the use of IE, the Bank's intention has been to promote more effective interventions. The introduction of the Development Effectiveness Matrix (DEM), the first Bankwide effort to promote more evaluable projects, has strongly incentivized the production and use of IEs. The DEM includes a series of monitoring and evaluation (M&E) requirements, and operations are rated according to their compliance with these requirements. This evolution has happened in the context of an overall cultural shift toward evaluation and more rigorous analysis, which were key components of the Ninth General Capital Increase and of the New Institutional Strategy. It is in this context that OVE assessed the production and use of IEs in IDB public sector operations, seeking to take stock and reflect on what this effort has brought to the Bank, what the cost has been, and what direction the Bank should take in the coming years.

The Bank maintains no central database, official registry, or budgetary classification of IEs; there is no institutional mechanism by which production of IEs can be easily determined. Furthermore, because IE can be produced in many different contexts, it is difficult to identify each IE by activities or registries.¹ OVE created a database to analyze each IE proposed by the Bank in loan documents and technical cooperation profiles, and gathered additional information through desk review and extensive interviews with IDB staff, other MDBs, clients and IE experts.

From 2006 and 2016 the Bank proposed 531 impact evaluations in loan documents and technical cooperation projects (TCs). Of these, 94 have been completed, 286 are ongoing (at different stages), and 151 have been cancelled for such reasons as project cancellation, political changes, and problems in design and implementation, including a lack of a well-defined methodology for IE at approval. IDB staff design 60% of IEs on average, a share that reached 80% in 2016. Among the IEs proposed in loan documents, about 34% have been motivated mostly by the DEM score, and these were more likely to be cancelled. From this it is possible to conclude that buy-in from governments and real interest in learning and using the results of the evaluation lead to more successful IEs.

The origination analysis shows a continuing lack of prioritization in selecting impact evaluations. A detailed review of the 20 current Sector Framework Documents (SFDs) shows that only around 30% both present an exhaustive literature review that identifies knowledge gaps and defines specific areas for analytical work, which could provide guidance for the prioritization of IEs.

The costs of IDB evaluations vary considerably but are broadly comparable with those of other organizations and with benchmarks found in the literature. The main driver of IE costs is the need to collect data: evaluations that require data collection cost on average US\$468,000, while the average cost of those that use administrative data is under US\$80,000. These amounts exclude the time of the Bank's specialists, but the data gathered for this evaluation indicate that data and consultants account for the bulk of the costs. On average the cost of an IE for a project is less than 0.5% of the loan amount.

Overall the Bank budgeted about US\$200 million for IEs during the 2006-2016 period, but about 25% of it was for evaluations that have been cancelled. Of the total, US\$152 million was paid for by operations, including loans and grants, and the remaining US\$54 million by TCs (with the latter not following any general strategy). When considering only the IEs in countries that did not have investment grants with IEs, IDB has spent approximately US\$0.61 from its ordinary capital and trust funds for every US\$1 spent by the countries to finance these IEs.

The Office of Evaluation and Oversight (OVE) identified many good practices across multilateral development banks (MDBs) for the financing of knowledge production. A transparent and strategic prioritization process allows MDBs to target knowledge gaps in specific sectors, facilitates the leveraging of resources, and reduces the costs of implementing and disseminating knowledge by fostering long-term relationships. Successful models for IE financing often involve active participation from knowledge end-users. Given that knowledge is a public good, cooperation across networks facilitates the production of IEs, as well as their relevance and visibility. Finally, a clearly defined research protocol has been used in most organizations, helping ensure publication of findings and attracting donors' participation.

The quality of the Bank's IEs seems to be increasing over time, but there is still room for improvement. OVE hired the CLEAR-LAC Center to assess the quality of completed IEs and of a sample of evaluation proposals. The criteria used by CLEAR followed international standards and included the relevance of the evaluation question, the appropriateness of the data used, the rigorousness of the method, and the robustness of the analysis. CLEAR specialists concluded that about 55% of completed IEs were satisfactory or partially satisfactory in quality. After the introduction of the DEM the quality decreased, as less traditional sectors started to do IEs as well, but it continued to increase afterwards. Similarly, the quality of the proposals shows a clear upward trend.

Regardless of their quality, IEs can be influential only if they are used, and OVE found that accessibility remains an issue. Among completed evaluations, 55 (58%) were not published in journals or in IDB working paper series. Many of them could not even be found in IDBDOCS, and OVE had to request them directly from the team leader. This shows that IE results are not easily accessible to the public. Even so, a citation analysis shows that the use of IDB IEs is increasing over time, helping IDB gain some recognition as a knowledge Bank.

Analysis of Bank documents suggests that evidence drawn from IEs is not often mentioned, though some clients seem to be using the IEs more directly. Clients had mixed views on the Bank's approach to IEs, but only 2 of the 25 interviews were highly negative. The other interviewees pointed to weaknesses in the process, but most were satisfied with the results—either because the IE helped ensure the sustainability or continuity of projects, or because it created capacity for them to continue doing this type of evaluation.

An important use of IE inside the Bank is to avoid ineffective programs. Although there were cases in which evaluations pointed to no results but the Bank proceeded with the same type of program anyway, IEs seem generally to have been used to avoid ineffective programs. This is likely to be one of the largest impacts of IDB's IEs, as the Bank helps governments avoid spending for programs that will not achieve the desired results.

OVE also found some evidence that well-justified IEs may help support project implementation. While the disbursement curves of projects with and without IE are almost the same during the first two years, projects with IE have faster disbursements afterwards and are completed three months earlier. Econometric analysis shows that these results are robust after controlling for several variables. While the exact channels through which IEs may influence project implementation are not clear, there is some evidence that support and training provided by researchers and field staff monitoring the project may play a role. And the financial implications of this better performance could potentially offset up to half the cost of the IE.

Reflecting the findings of this evaluation, OVE has the following recommendations to improve the production and use of IEs in IDB:

1. Be more strategic in the selection of IEs by undertaking or supporting IEs only if they have client commitment (or a clear strategy to engage the client to develop such commitment), a well-identified knowledge gap, and a feasible timeframe. This evaluation has shown the importance of having the client on board when conducting an IE. Client commitment also helps to ensure the use of IE results at completion.
2. Ensure that SFDs identify knowledge gaps to help guide the Bank's IE work. SFDs have done this to varying extents to date, and OVE recommends that they do so more consistently and thoroughly going forward.
3. Revise the DEM to reduce the incentives to over-propose IEs. Every project should count on a solid monitoring and evaluation system, but not all of them need to have an impact evaluation. The DEM has been a useful tool to increase projects' evaluability and to promote more understanding of the importance of doing more rigorous evaluations. It has also increased the incentives to propose IEs, leading to proposals that were not well-thought out or warranted.
4. Develop a transparent funding mechanism, aligning the interests of clients and the Bank. When an IE is deemed relevant and feasible for a loan, it should be incorporated into the M&E system funded by that loan to the extent that it will be informative for the country and for the preparation of the Project Completion Report (PCR). If there are longer-term impacts that can be measured only after the project closes, post-project evaluation work should be funded by Bank-managed resources. In this way clients who are committed to IEs and put in place proper M&E arrangements could benefit from IDB resources to complete the evaluation after the project closes. At the same time, the Bank would commit its own-managed resources only to those longer-term evaluations where there was client commitment, good early management (with proper baselines and follow-up), and an expectation of valuable information on longer-term program results.
5. Strengthen systems for quality control. The Bank already has a number of mechanisms in place for ex-ante quality control. However, as shown in this evaluation, not all IEs are completed with high quality. To avoid the reputational risk of publishing and disseminating low quality work, the Bank needs to develop a system for quality control beyond the IEs published as Working Papers or Technical Notes.

6. Strengthen and move towards centralization of the Bank's systems to track all impact evaluations being conducted—from design to completion—and their costs, and to make high quality IEs and their documentation, the data used (when possible), and their lessons available to the public. SPD and KNL have made progress in these directions, as noted in this evaluation, and the system should be completed.
7. Develop a more formal mechanism to promote partnerships to increase the quality, relevance, and visibility of IDB work, while potentially reducing costs. This evaluation found little partnership in the actual production of IEs. Based on what has been observed as good practice in IDB and other organizations, the Bank could usefully increase its partnerships with academia as well as its collaboration with other organizations in IE production, dissemination, and information-sharing. SPD has been promoting this, and the effort should be strengthened.

¹ IE can be produced in four different ways: (i) as pure research product, (ii) as input during the preparation of projects, (iii) as part of stand-alone sector work, and (iv) as ex-post product that is included as part of the Bank's normal project evaluation cycle.



1

Since the mid-2000s the Bank has been involved in the production of many IEs. Between 2006 and 2016 more than 400 operations proposed at least one IE.

© IDB

1 Introduction

The Inter-American Development Bank (IDB, or the Bank) has growing experience in doing impact evaluations (see Box 1.1 on Impact Evaluation definition), and this report takes stock of what has been done so far. Since the mid-2000s the Bank has been involved in the production of many IEs. Between 2006 and 2016 more than 400 operations (including loans and technical cooperation projects) proposed at least one IE, and almost 100 IEs have been completed. It is now time to reflect on what this effort has brought to the Bank, what the cost has been, and what direction the Bank should take from here. The analysis is divided into two parts: 1) The assessment of IE production: checking on what has been proposed and their status, the production process, the cost of the IEs proposed, and the quality of completed IEs and of proposals; 2) The assessment of the use and influence of IE: considering the internal use in defining IDBs agenda and project design, the use of the IEs among our clients, and the general use, as well as the potential impact on project implementation.

Impact evaluations (IEs) have long been used to estimate the causal effect of an intervention, particularly in the medical literature. Since the early 2000s they have been extensively used to assess the effects of public policies as well.¹ The purpose of an IE—particularly when done through a randomized control trial (RCT)—is to isolate and measure the effect of an intervention. IE generates a better understanding about which policies and programs work and which do not. Methodological advances and greater availability of high-quality data have allowed for an increase in the types of questions that can be answered through IEs, mainly during the last decade. In a recent review of IEs, Dhaliwal and Tulloch (2012) argue that evidence from IE, when it comes from methodologically rigorous and independent analysis, can influence policymaking

and provide evidence that is easy to understand and actionable. In addition, IEs can inform the scaling-up of public policies and ensure the sustainability of projects with proven effectiveness.

Box 1.1 - What is an Impact Evaluation?

According to the OECD, an impact evaluation is “an assessment of how the intervention being evaluated affects outcomes, whether these effects are intended or unintended (OECD, 2006). The analysis requires a counterfactual of what those outcomes would have been in the absence of the intervention.” Gertler et.al. (2016) describe impact evaluations as a “particular type of evaluation that seeks to answer a specific cause-and-effect question: What is the impact (or causal effect) of a program on an outcome of interest?” The main concept is to differentiate the real effect that can be attributed to any of three different cases: program, program modality or design innovation.

Many different methods can be used to address the cause and effect question, but in each case the estimation of the causal effect needs to find an appropriate counterfactual: the situation of the outcome(s) of interest in the absence of the policy or intervention for a group with similar characteristics. The selection of the impact evaluation methodology must be closely related to the operational characteristics of the intervention, the rules for its implementation and how the data is collected. When these three elements are clear, the causal effect is relatively easy to measure [NONIE (2009) and Gertler et. al. (2016)]. Some relevant terminology is defined below:

- ***Multivariate regression.*** Individuals who received treatment are compared with those who did not, and other factors that might explain differences in the outcomes are “controlled” for.
- ***Differences in Difference.*** The improvement (change) over time of program participants is measured relative to the improvement (change) of non-participants.
- ***Statistical Matching.*** Individuals in a control group are compared to similar individuals in an experimental group.
- ***Regression Discontinuity Design.*** Individuals are ranked based on specific, measurable criteria. There is some cutoff that determines whether an individual is eligible to participate. Participants are then compared to non-participants and the eligibility criterion is controlled for.
- ***Instrumental Variables.*** Participation is predicted by an incidental (almost random) factor, or “instrumental” variable, that is uncorrelated with the outcome, other than the fact that it predicts participation (and participation affects the outcome).
- ***Experimental Evaluation, or Randomized Control Trial (RCT).*** Participants are randomly assigned to experimental and control groups to measure a causal relationship between two variables.

Development agencies are in a privileged position to encourage a better understanding of policies and programs, which improves their own ability to sponsor effective interventions for development. More than a decade ago a Center for Global Development report posed a provocative question— “When will we ever learn?”²—calling on development organizations to use IE. The concept behind this report was not that every development project should include an IE, but rather that there were many opportunities to learn from the projects being implemented by multilateral organizations. The report also suggested that, given the nature of knowledge as a public good, coordination among all organizations in this effort would enhance this learning process. Prodded by this evidence-based agenda, development agencies have implemented many initiatives to strengthen their evaluation capacities and resources, including investment to finance ex-post evaluations and RCTs—thus strongly contributing to the growth in the production of IEs (Box 1.2).

Although IEs are widely used to measure the effectiveness of foreign aid, their usefulness—particularly for RCTs—has been challenged in recent years. There are four main criticisms of IEs.

- Deaton (2009) argues that “experiments have no special ability to produce more credible knowledge than other methods, and that actual experiments are frequently subject to practical problems that undermine any claims to statistical or epistemic superiority.” Regardless of the method used, the author maintains, the important thing is to have a clear theory of the mechanisms behind the intervention being assessed, which in the last instance is what should be evaluated.
- Many researchers have responded that one IE alone is not enough to inform policy.³ They suggest implementing other types of studies in addition to IEs for building knowledge. IEs alone can often answer questions on what works, but many times they may fail to address the relevant question of why it works. When followed by process tracing and qualitative evaluations, an IE can be much more informative, particularly for changing policy.
- A third major source of concern relates to external validity. Many scholars have argued that the results of IEs are very much context-dependent, hindering their ability to be informative in other contexts (Deaton, 2010; Pritchett and Sandefur, 2013). Vivaldi (2015), after analyzing more than 20 diverse interventions from numerous studies and showing a large heterogeneity across their results, argues that generalizability can be achieved by finding differences in interventions and selected populations. She says that working in this direction would help improve the predictability of IE results.
- Related to the external validity discussion, a more recent concern about IEs is the focus on the average effects of a program or policy, without looking at its distributional effects.⁴ Many scholars have developed different methodologies to measure these effects.

In sum, the trend of academic thought and day-to-day practice has been to seek greater focus on understanding the mechanisms behind interventions rather than on just measuring their results, and on combining IEs with other types of evaluations to build comprehensive knowledge of what works.

Box 1.2 - IE in Other Organizations

With the creation of the Development Impact Evaluation Initiative (DIME) in 2005, the World Bank began a more systematic use of IEs to measure results, promote learning, and inform institutional strategies. The initiative contributed to increasing the number of completed IEs at the World Bank from 28 in 2005 to 215 in 2015. At the programmatic level, DIME facilitates the process of setting the research agenda, managing research efforts, and carrying out dissemination, policy outreach, and fund-raising activities.

The Millennium Challenge Corporation (MCC) has also provided funding for IE and revised its mandates to require these evaluations. Created by the U.S. Congress in 2004, the MCC has sought to deliver U.S. foreign assistance by focusing on good policies, country ownership, and results. It pushed IEs heavily during its first six years as part of its results framework; in 2011 more than 60% of its project evaluations were IEs. Today the organization is more strategic in proposing IEs, using a revised feasibility framework. Still, in 2016 the percentage of projects with an IE was close to 40%.

Academia has also established networks that support the use of IEs to study development policy. Among the most prominent is the Abdul Latif Jameel Poverty Action Lab (J-PAL), which was founded in 2003 in the Department of Economics of the Massachusetts Institute of Technology and works with over 20 governments in 11 countries to institutionalize the use of evidence in decision-making. Since its creation it has produced 842 ongoing and completed IEs through a network of more than 140 affiliated professors at 49 universities around the world. The organization has independent regional offices hosted at leading universities in Africa, Europe, Latin America and the Caribbean, North America, South Asia, and Southeast Asia.

The Bank has been at the forefront of the production and use of IE in Latin America and the Caribbean, particularly since its involvement with the evaluation of PROGRESA in Mexico. IDB provided funding for and closely followed the International Food Policy Research Institute's IE of PROGRESA,⁵ and a few years later included the use of IE as best practice in similar projects.⁶ This early IDB effort was focused in the social sectors because of the availability of data and familiarity with evaluation methods in many countries. In other sectors the process of developing a culture of evaluation based on IE has been slower.

Beginning in the early 2000s, the IDB's Office of Evaluation and Oversight (OVE) started to push IDB to focus more on evaluation. In 2002 OVE published its first assessment of the Bank's evaluability, setting minimum standards for a project to

be evaluable and showing that for most projects IDB was missing the opportunity to deepen its knowledge on effectiveness.⁷ OVE's 2005 Evaluability Report recommended that evaluability standards be introduced as a criterion for project approval. This recommendation was adopted and included in the institution's mandate as part of the Ninth General Capital Increase (GCI-9). While OVE was pushing for greater evaluability to be built into projects at the design stage, it was also starting its own ex-post IE agenda. In 2004, the Board of Directors gave OVE a mandate to undertake systematic IEs of IDB projects.⁸

Only after the Realignment and the beginning of the Knowledge Bank agenda were conditions created for the promotion of IEs by operational units in the Bank. Two key changes accelerated this process. The first was the change in staff profile, as the Realignment allowed the Bank to increase its internal capacity for research, including IEs, and the number of specialists with more academic training increased substantially.⁹ The second was the creation of the Office of Strategic Planning and Development Effectiveness (SPD). As the Realignment document stated, "The creation of the Office of Strategic Planning and Development Effectiveness is an unequivocal signal of the cultural change embodied in the realignment"; part of SPD's mandate was to help ensure the Bank's development effectiveness by "provid[ing] technical expertise to the project cycle."¹⁰

In favoring the use of IE, the Bank's intention has been to promote more effective interventions. The Development Effectiveness Framework (DEF),¹¹ prepared by SPD and approved in 2008, internalized the idea that the Bank would improve its development effectiveness through a robust understanding of which interventions work and which do not. The DEF better defined Bank units' roles and responsibilities regarding development effectiveness and introduced the Development Effectiveness Matrix (DEM) to assess operations' ex-ante evaluability. The DEM introduced a series of monitoring and evaluation (M&E) requirements, with operations rated on their ex-ante compliance with these requirements. The system explicitly recognized the importance of estimating the causal effectiveness of operations by giving higher ratings to those with evaluation plans that included an IE. Thus since 2009 the DEM has strongly incentivized the production of IEs.

The cultural shift toward generating evidence-based interventions was also a key component of the GCI-9 and of the IDB's New Institutional Strategy. Both documents aimed to strengthen the institutional framework to deliver knowledge products to better serve the region. The IDB update to the Institutional Strategy 2010-2020 states:

"To promote the adaptation of successful development approaches, it is important to strengthen the Bank's capacity to learn and to disseminate what it has learned. This calls for increasing the efforts to generate knowledge of what works and what does not, based on rigorous evidence—which, in turn, requires

continuing to evaluate projects throughout their life cycle. Furthermore, so that innovation can permeate throughout the institution, the IDB needs to create a new culture where collaboration, diversity, experimentation, and practice are the main ingredients, resulting in better solutions, processes, and products.”¹²

In this context, OVE assessed the production and use of IEs in IDB public sector operations.¹³ The analysis focused on the public sector because almost all of the effort has been undertaken there to date. OVE collected detailed data on all loans and TCs that included at least one impact evaluation and were approved between 2006 and 2016 (Box 2.1). Given that the DEF was approved in 2008 and the DEM launched in 2009, OVE decided to include a few prior years as a baseline. OVE also complemented its analysis of IDB IE production by hiring the CLEAR Center for Latin America (through a competitive selection process) to do a systematic assessment of the quality of the majority of completed IEs and a sample of proposals of ongoing IEs. Additionally, OVE conducted a citation analysis of published IEs and interviewed a sample of IDB clients to gather information on client’s view and use of IEs. In total OVE conducted more than 200 interviews with bank staff working with IE or producing IEs, sector managers, division chiefs, REAs, and representatives of other organizations working with IEs. Table 1.1 summarizes the methods used.

TABLE I.1 - EVALUATIVE QUESTIONS AND MEANS OF VERIFICATION

Topic	Evaluative Question	Means to address the questions
Analysis of production	How many IEs have been proposed?	<ul style="list-style-type: none"> Construction of database of all IEs proposed in loan documents and TCs – based on desk review and interviews with bank staff in charge
	What is the status of proposed IEs?	
	What was the origination, design and implementation process?	<ul style="list-style-type: none"> Interviews with Bank specialists, division chiefs, sector managers, and clients
	How much do IEs cost?	
	What are best practices in financing impact evaluations and knowledge products in general?	<ul style="list-style-type: none"> Interviews with Bank staff involved in the production of IEs and representative of other organizations working with IEs.
	What is the quality of IDB IEs and how has it evolved over time?	<ul style="list-style-type: none"> CLEAR assessment based on standard criteria of 84 completed evaluations and 59 proposals of ongoing ones.
Analysis of use and influence	How is the knowledge produced by IEs used inside the Bank?	<ul style="list-style-type: none"> Desk review of country strategies, technical notes, and Sector Framework Documents. Citation analysis on loans approved in 2007 and 2016. Interviews with division chiefs, sector managers, and REAs. Interviews with Bank staff working with impact evaluations
	How is knowledge produced by IEs used by clients?	<ul style="list-style-type: none"> Interview of a sample of 25 Bank clients in projects with IEs
	How are IEs used in general?	<ul style="list-style-type: none"> Citation analysis of all published IEs between 2006 and 2016.
	Do IEs help in delivering projects?	<ul style="list-style-type: none"> Econometric analysis of project disbursements.

In this document, OVE focuses on the IEs proposed in loan documents and technical cooperation projects (TCs).

© IDB

2 The Production of Impact Evaluations in IDB

The Bank maintains no central database, official registry, or budgetary classification for IE; thus there is no institutional mechanism by which production of IEs can be easily determined. Furthermore, because IE can be produced in many different contexts, it is difficult to identify IEs through activities or registries.¹⁴ In this document, OVE focuses on the IEs proposed in loan documents and technical cooperation projects (TCs). Box 2.1 summarizes the methodology used to identify the universe of IEs and its limitations (see Annex I).

A. ANALYSIS OF PRODUCTION

1. How many IEs have been proposed and what is their status?

The institutional changes implemented in the IDB, particularly the DEM, resulted in a sharp increase in the number of loans and TCs that included an IE. From 2006 to 2008, an average of 8% of loans and 2% of TCs per year included an IE. This increased to an average of 28% and 3%, respectively, from 2009 to 2016. From 2006 to 2008 an average of 13 IEs were proposed per year, while from 2009 to 2016 the annual average was 61. The number of proposals increased sharply from the introduction of the DEM until 2013 and declined somewhat thereafter (Figures 2.1 and 2.2).

A total of 531 IEs were included in loans and TCs approved between 2006 and 2016. Of the proposed IEs, 94 (18%) have been completed, 151 (28%) have been cancelled,¹⁵ and 286 (54%) remain in the pipeline. As shown in Table 2.1, SCL has proposed the most IEs (47%), followed by IFD (28%) and CSD (14%). INT

has completed 15 IEs financed by ESWs, which are not reflected in Table 2.1. The stage of ongoing IEs varies from design (40%) to final data analysis (11%), with most expected to be completed within the next 5 years.¹⁶

Box 2.1 – Methodology

Because of the dispersion of documents and multiple sources, OVE adopted a four-step strategy to maximize its chances of successfully recognizing every IE executed or funded by the Bank.

- Initial screening of loan documents and TC profiles. A lexical analysis of all loan documents registered in IDB-SEC and the most recent version of TC profiles published in IDBDOCS between 2006 and 2016, followed by a thorough desk review of these documents.
- Interviews with team leaders and division chiefs. Structured interviews with the team leaders or specialist in charge of the IEs in each division to collect data not contained in the loan proposals, M&E annexes, or TC profiles.
- Validation process with SPD. A comparison of OVE's database with all IEs registered in SPD's database.
- Database processing. The creation of a database containing all relevant information for each identified IE. The analysis in this report draws on this database.

OVE identified 414 operations (122 TCs and 292 loan operations) that included at least one IE. Since many loans and TCs had multiple IEs, OVE's database has information on 531 IEs proposed between 2006 and 2016. Some IEs were designed and conducted after the operations' approval, and OVE had no mechanism to track these cases.

OVE carried out a complementary effort to identify IDB production of IEs and compare it to global trends. Given that IEs can also be produced with resources from economic and sector work and from the administrative budget, OVE's approach did not identify all IDB IEs done or in process. So, it constructed a secondary database for IEs published, including information about individual studies (title, year, author, and department) and links to the studies themselves. IDB publications based on IEs increased sharply since 2009 and seem to follow the same pattern found by Cameron, Mishra, and Brown (2016) regarding growth, sector distribution, and methodologies. Although the IEs were initially heavily concentrated in the social sectors, over time other sectors have created more IEs.

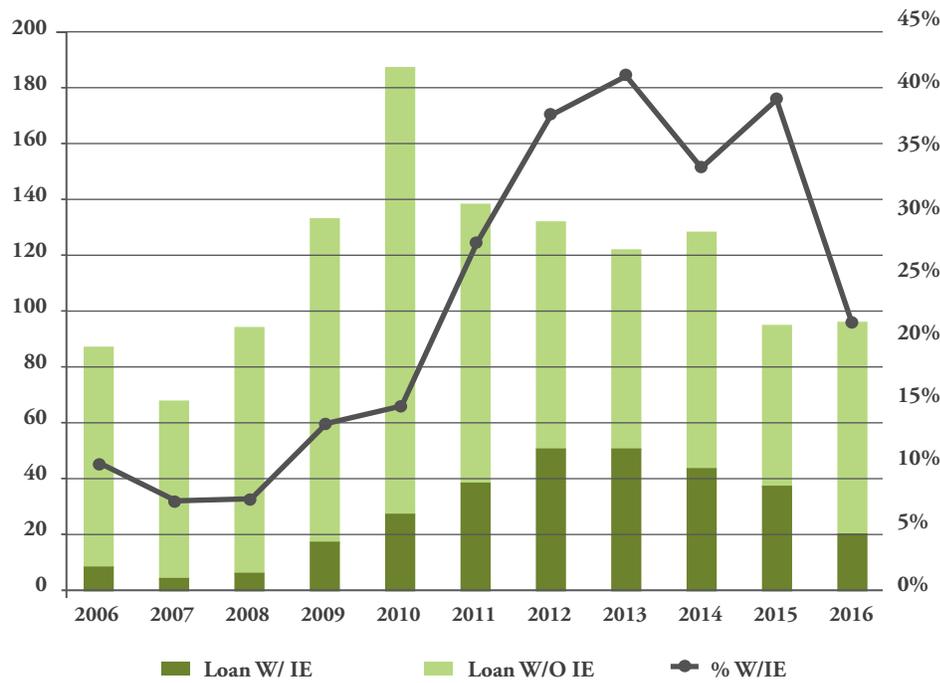


FIGURE 2.1
IDB Loans with IE

Source: OVE, IE Database 2017.

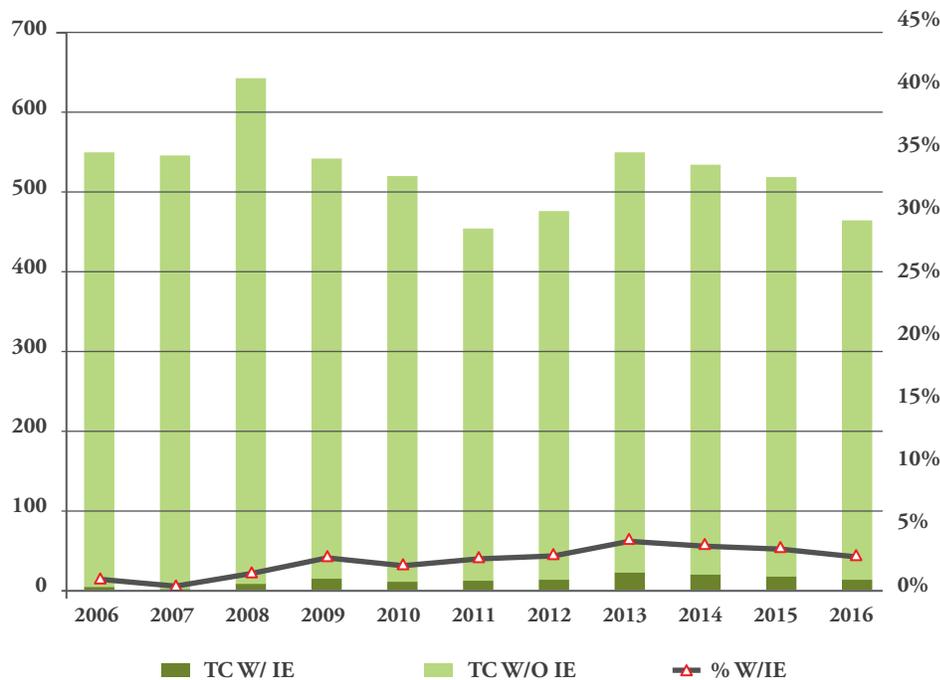


FIGURE 2.2
IDB TCs with IE

Source: OVE, IE Database 2017.

TABLE 2.1 - DISTRIBUTION OF IE BY DEPARTMENT, DIVISION, AND STATUS

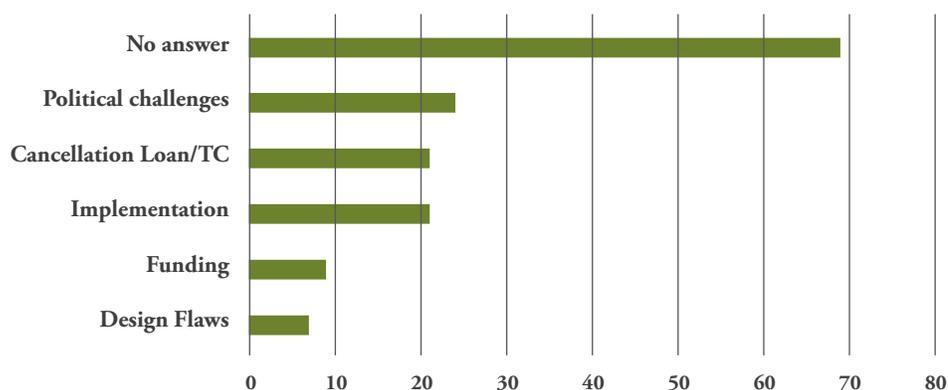
Department	Division	Loans				TCs			
		Completed	Ongoing	Cancelled	Total	Completed	Ongoing	Cancelled	Total
CSD	CCS	0	1	1	2	0	1	0	1
	RND	6	33	9	48	1	2	2	5
	HUD	1	9	9	19	0	1	0	1
IFD	CMF	3	8	10	21	0	0	3	3
	CTI	5	26	8	39	2	3	3	8
	FMM	0	5	2	7	0	0	0	0
	ICS	3	29	12	44	7	13	8	28
INE	ENE	0	3	1	4	0	0	0	0
	TSP	1	12	0	13	0	1	0	1
	WSA	0	5	3	8	0	1	1	2
INT	TIN	0	20	2	22	0	0	1	1
SCL	EDU	9	19	21	49	17	11	6	34
	GDI	1	2	0	3	6	7	5	18
	LMK	1	5	7	13	11	3	4	18
	SPH	11	58	25	94	9	8	8	25
Total		41	235	110	386	53	51	41	145

Sources: OVE. IE Database 2017; OVEDA (last column); Includes IEs executed by VPC (CAN and CCB), RES and SCL front office, classified by topic and sector.

OVE’s interviews with team leaders¹⁷ indicated that the main reasons for IE cancellation, when the loans and TCs were not themselves cancelled, appear to have been political challenges (30%) and implementation and design issues (34%) (Figure 2.3). IEs proposed mainly to increase DEM scores--rather than because of a genuine interest in learning on the part of IDB or the client--were significantly more likely to be cancelled.¹⁸ Having a well-defined IE method is associated with a lower probability of cancellation, especially when the method is an RCT. In fact, if an RCT starts but fails for any reason, the teams are more likely to find other ways to conduct an alternative IE.¹⁹ The probability of cancellation does not vary significantly by country or by cost of the IE.

FIGURE 2.3
Main Reason for Cancellation of IE

Source: OVE. IE Database 2017



IEs have been unevenly distributed among country departments and sectors. With regard to country distribution, CID has the highest share at 35% and CCB the lowest at 7% (see Annex II). With regard to sector distribution, completed IEs are heavily concentrated in earlier projects (Figure 2.4) in the social sectors (Figure 2.5). Few IEs in other sectors have been completed, though their numbers (particularly in IFD) have increased.²⁰ Some IEs have been financed through ESW (notably in INT) are not captured here.²¹ The social sectors have traditionally conducted more IEs in part because it is easier to observe who the beneficiaries are and define a clear methodology to assess the effect of an intervention.²² Yet sectors in which it is extremely difficult to do IEs have made increasing efforts to innovate using different methodologies.²³ In addition, OVE noted some differences in the reasons for IE cancellation, with the main drivers for SCL, INE, and INT being political factors and implementation issues, while for CSD and IFD they were design issues and the cancellation of underlying operations.

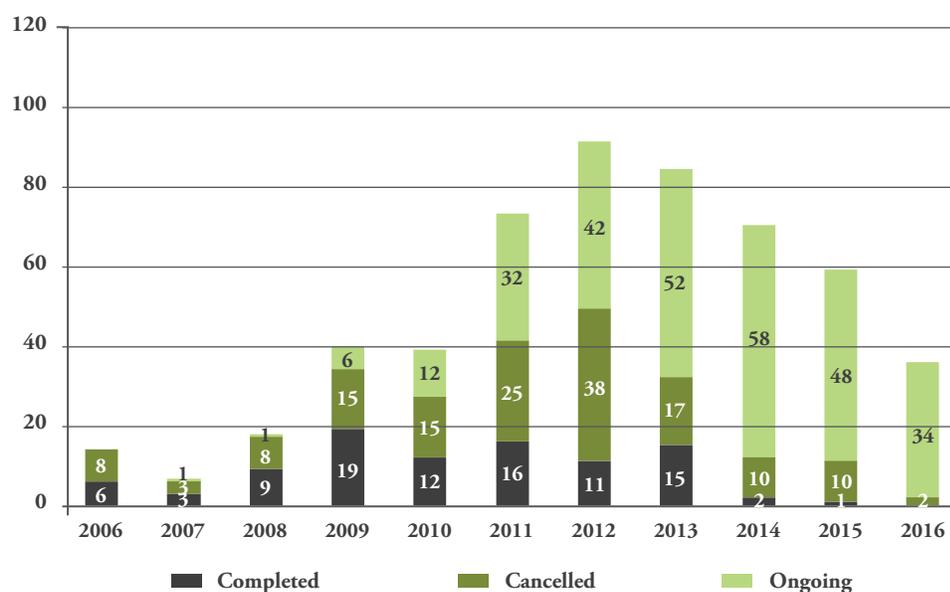


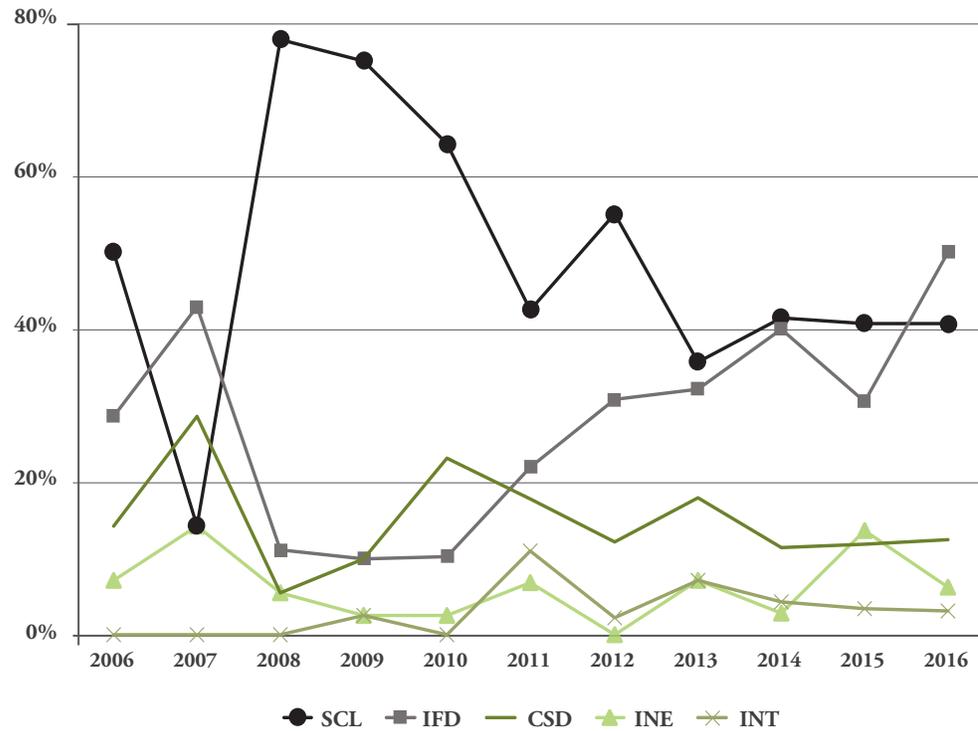
FIGURE 2.4
Status of IE by Year of Approval

Source: OVE. IE Database 2017.

The Bank has diversified not only the sectors doing IEs, but also the type of methodologies used. The number of evaluations proposed with an undefined methodology has decreased (Figure 2.6), and in 2016 all projects had a clear methodology at approval. This is an indication that more thought is being given to the design of the IEs and the causal mechanism that is being tested in the projects. To the extent that evaluations with a well-defined methodology are less likely to be cancelled (as noted above), there should be fewer cancellations over time. The overall absolute and relative numbers of RCTs have declined, and the numbers of quasi-experimental IEs have increased correspondingly. Though data collection is still needed, since 2011 more evaluations have used administrative data.

FIGURE 2.5
Distribution of IE
by Department

Source: OVE, IE
Database 2017.



The trends reported above have some important implications for the Bank. First, many evaluations have been cancelled,²⁴ which in itself is not necessarily a bad outcome; in some instances, a different type of evaluation could be more informative than pursuing an IE that does not meet the conditions to be completed and used. However, an important problem may arise when the IE was accompanied by generally poor M&E planning. If an IE is the only source of information, and no suitable M&E system is in place, its cancellation will mean that the Bank has no information on the results of the project.²⁵ Second, the Bank has generated a large portfolio of ongoing IEs. As these evaluations are completed, it will become more and more imperative for the Bank to have a system in place not only to track what is being produced, as done through BRIK for publications, but also to absorb the knowledge being generated, with a strategy to organize and disseminate the findings. Finally, the reduction in the number of IEs and evaluations being proposed without methodology suggest that the Bank may be beginning to move toward more strategic use of IEs.

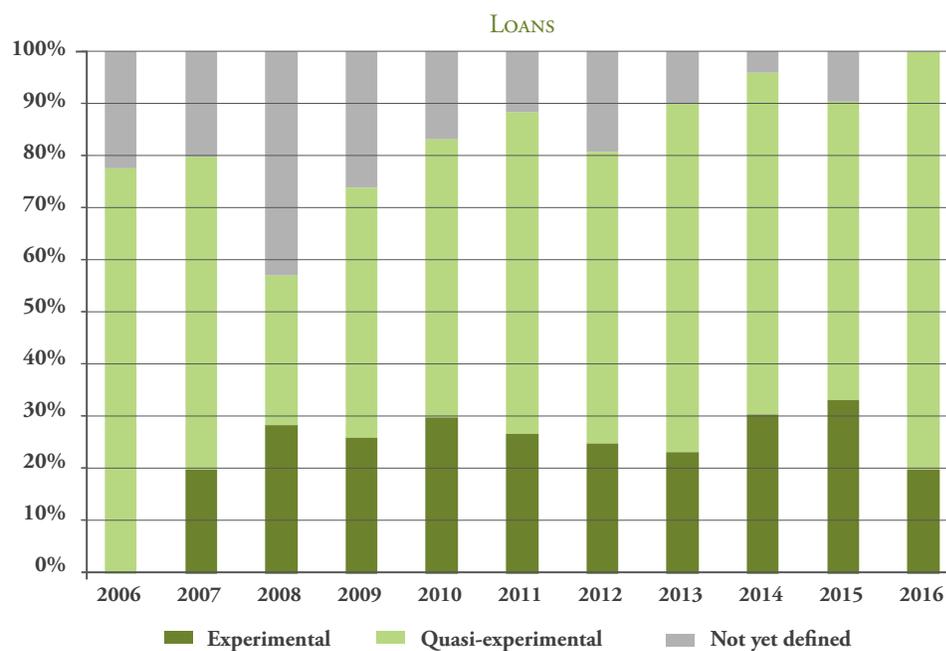
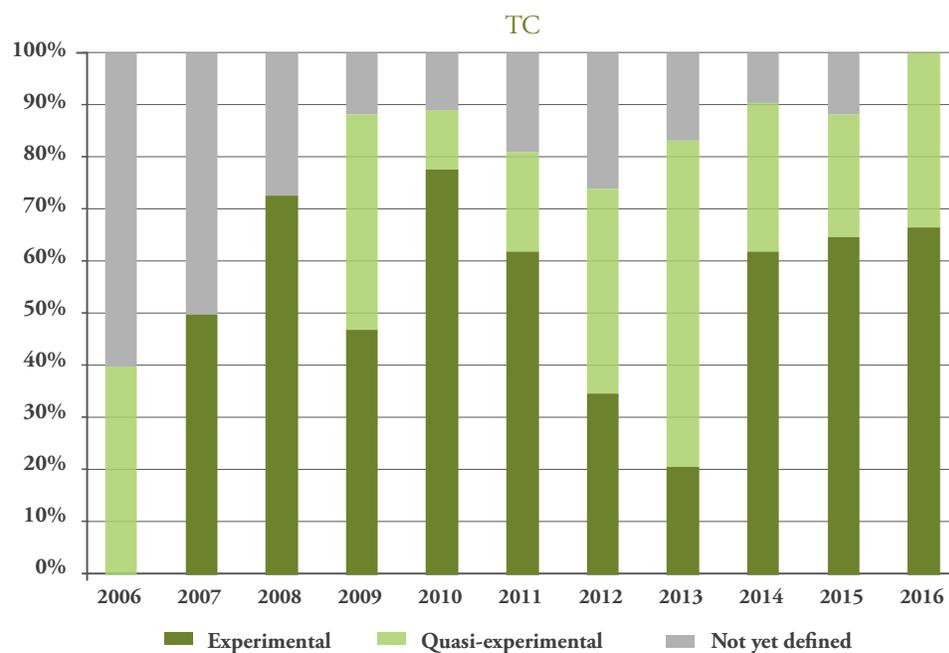


FIGURE 2.6
Distribution of IE by Methodology and Year

Source: OVE, IE Database 2017.



2. How have the processes for origination, design and implementation worked?

The main problem with the steep increase in IE production is the lack of structure behind the generation process. The Bank has set strong incentives to propose IEs, but has not yet developed a mechanism to propose them in a strategic way. In 2012 the Bank put in

place a new regulatory framework to govern sector interventions, introducing the Sector Framework Document (SFD) “to provide meaningful guidance to project teams and provide a clear sense of what the Bank seeks to accomplish in each sector” (GN-2670-1) The SFDs have a well-defined structure that highlights the importance of evidence and development effectiveness, and they are supposed to present a thorough review of international evidence concerning the main challenges of the region. Although SFDs should identify key knowledge work by sector, different sectors have varying degrees of success in proactively identifying knowledge gaps and translating them into specific lines of action. A detailed review of the 20 current SFDs (Annex VII) shows that only around 30% present a thorough literature review that identifies knowledge gaps to the extent needed to provide guidance for the prioritization of IEs. More than half of these SFDs are in the social sectors, which is not surprising given the maturity of IE production in education, social protection, health, and gender.

As the analysis above suggests, the strong incentives to do IE seem to have caused some level of “over-proposing,” and many IEs do not meet the criteria to be relevant or even feasible. IEs are not easy to produce, given the time, resources, and political will required for success.²⁶ It is understandable that some will be cancelled during project implementation. However, as noted earlier, an IE whose proposal is motivated mainly by the DEM is more likely to be cancelled, as real interest and buy-in from the government and project team may not exist. Among loans, about 34% of proposed IEs were motivated by the DEM score only (Figure 2.7). In 17% of cancellations, the specialist reported that the conditions to implement an IE were not in place, and eventually implementation became impossible.

A large share of the IEs reviewed—46% of those in loans and 66% in TCs--were Bank-driven (Figures 2.7 and 2.8). Only 14% of IEs in loans and 20% of IEs in TCs originated from the interests of clients. Interviewees noted that many IEs did not address questions that were relevant to clients and were often presented to clients as a Bank requirement. Thus, governments have often perceived IEs as a tax rather than a valuable product or service in support of project implementation.

When asked about the ownership of IEs, most staff were unsure. A common view was that the government owned IEs paid by loans, while the Bank owned IE paid by TCs (Figure 2.9), which is understandable, particularly for Bank-managed TCs. However, for about 30% of IEs in loans and 10% in TCs the team leader was unsure about ownership, and thus about how to proceed in consulting with clients on making results public. Clients experience the same confusion: many times they do not know what to do with the evaluation beyond internal use, even in cases in which results may be relevant to others.²⁷

IDB staff are centrally involved in the design of the IEs. Bank staff design 60% of IEs on average, and that share reached 80% in 2016 (Figure 2.10). There are no significant differences in this share across sectors. This is in part a response to the Bank’s

increasing emphasis on quality of design and the recognition that not all government agencies are equally equipped with capacity to do IEs. Once the project is approved, the implementation of the IE becomes the responsibility of the client, which usually (in 70% of cases) hires consultants to do the work, particularly for loans.²⁸ The Bank often supports governments in designing terms of reference for hiring consultants and helping to oversee their work. Some clients view this support as a positive service the Bank is offering, though others are not interested, perceiving the IE as another requirement of the Bank, and leave Bank staff to do it alone.

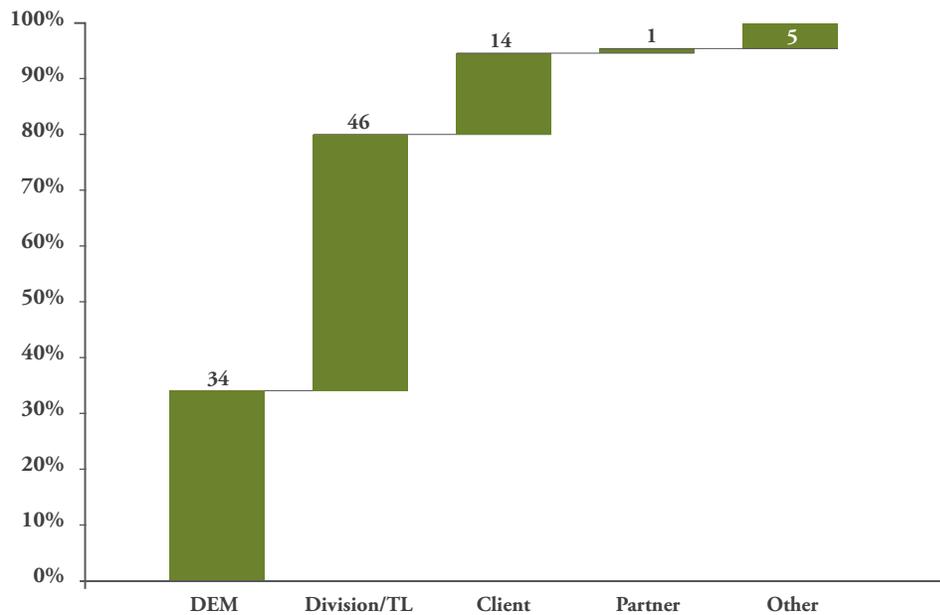


FIGURE 2.7
Origination of IE in Loans (%)

Source: OVE. IE Database 2017.

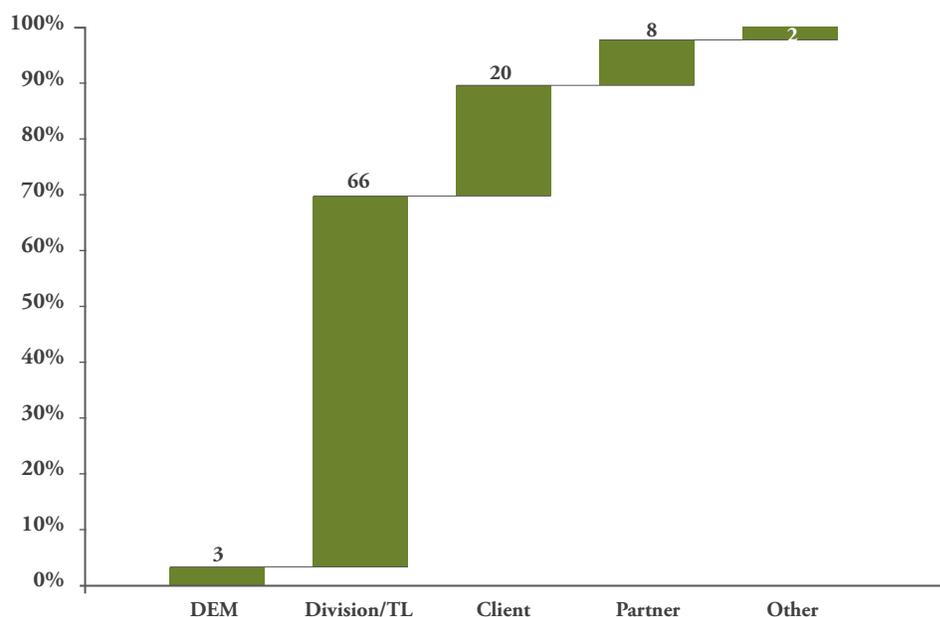


FIGURE 2.8
Origination of IE in TCs (%)

Source: OVE. IE Database 2017.

FIGURE 2.9
Team leader perception
of ownership of the IE
by Source of Budget

Source: OVE. IE Database
2017.

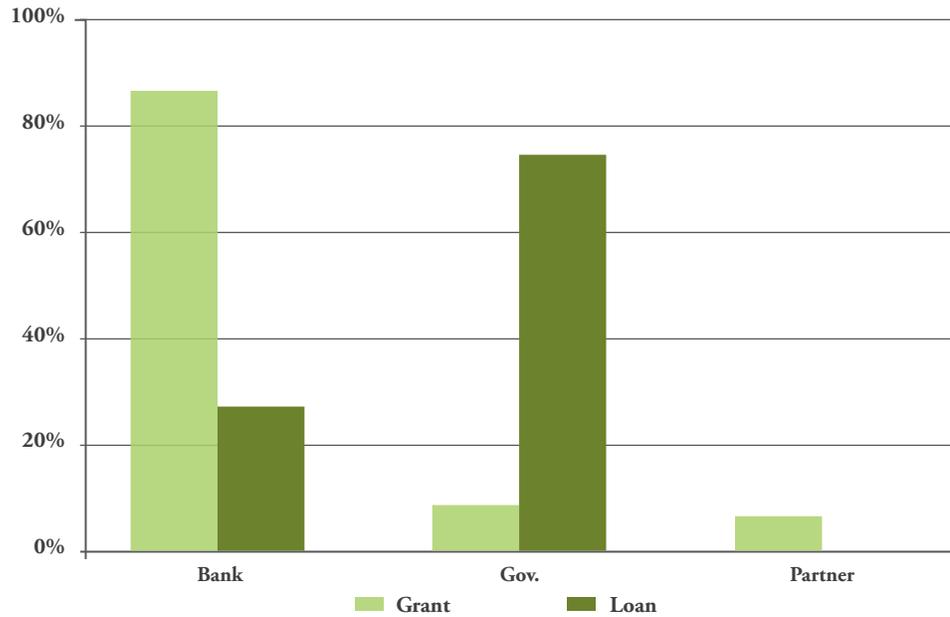
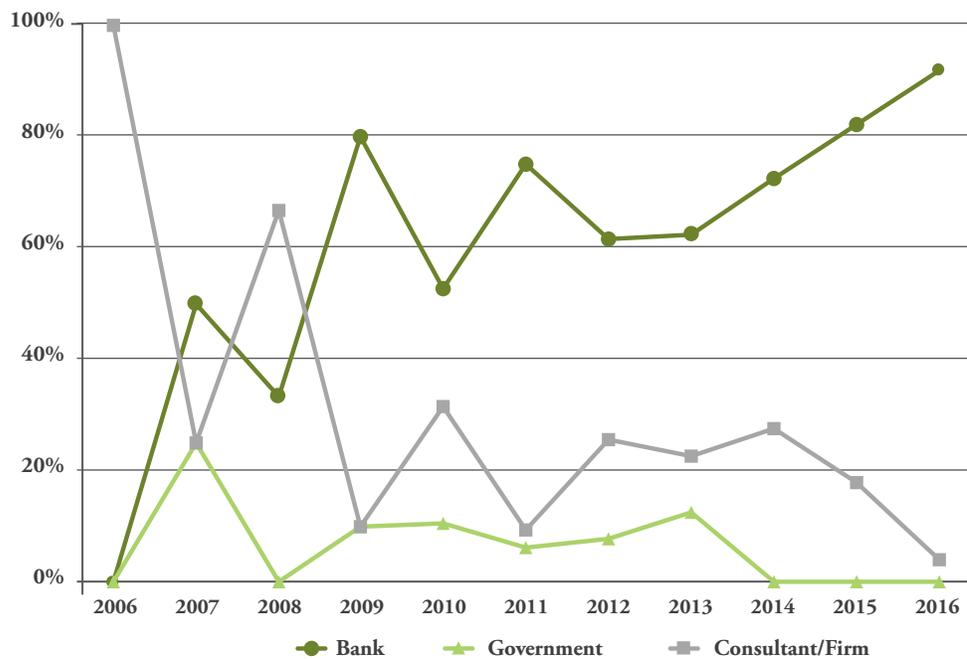


FIGURE 2.10
Design of IE by Year

Source: OVE. IE
Database 2017.



SPD has had a key role, through both quality control of the DEM for operations and more direct involvement in about 25% of the IEs. Bank staff indicated that they received substantive support from SPD in 35% and 46% of the IEs included in TCs and loans, respectively (Figure 2.11). SPD's support has been focused more on sectors with less expertise in IE, with INE receiving the most support (Figure 2.12). SPD is currently directly supporting about 70 ongoing IEs. Beyond ensuring minimum quality through the review process, SPD is participating more actively in design and sometimes in implementation.

The Bank has relied heavily on individual consultants and consulting firms in producing IEs, though few partnerships have been set up to promote IE (Figure 2.13).²⁹ Only 23% of the ongoing IEs are being conducted in partnership with an external institution, either for their design or implementation—a substantial decline from the 37% of completed IEs that benefitted from external collaboration.³⁰ Of cancelled IEs, only 10% were done in collaboration with other institutions. Of the total amount of IEs, only 1% of IEs have been or are being conducted entirely by partners, with no IDB direct involvement. This represents a missed opportunity to expand the scope of IEs, financial alternatives, and the dissemination of results. Other institutions have found it more attractive to collaborate with external partners.³¹

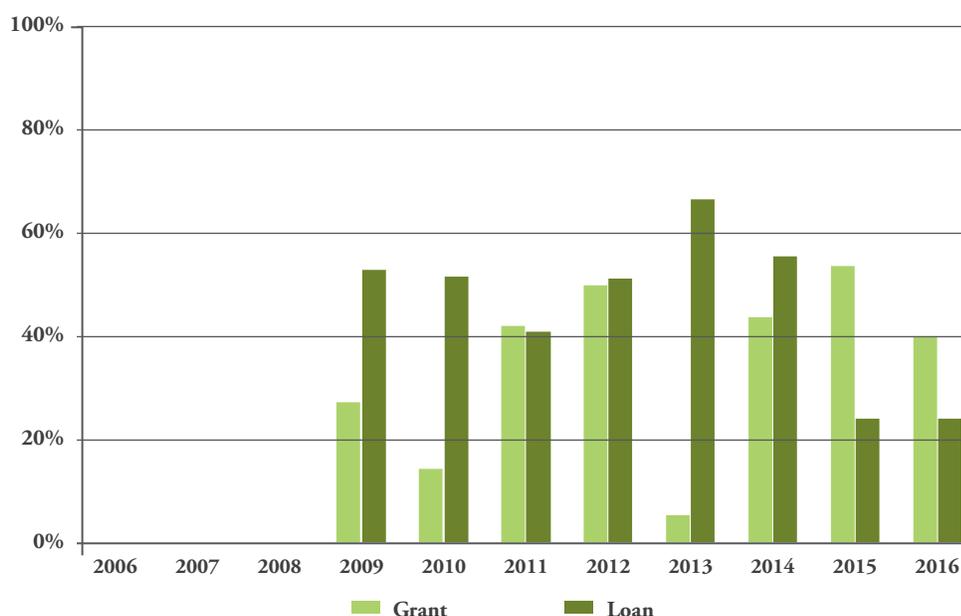


FIGURE 2.11

SPD Support by Year

Source: OVE. IE Database 2017.

Note: Grants includes TCs and investment grants

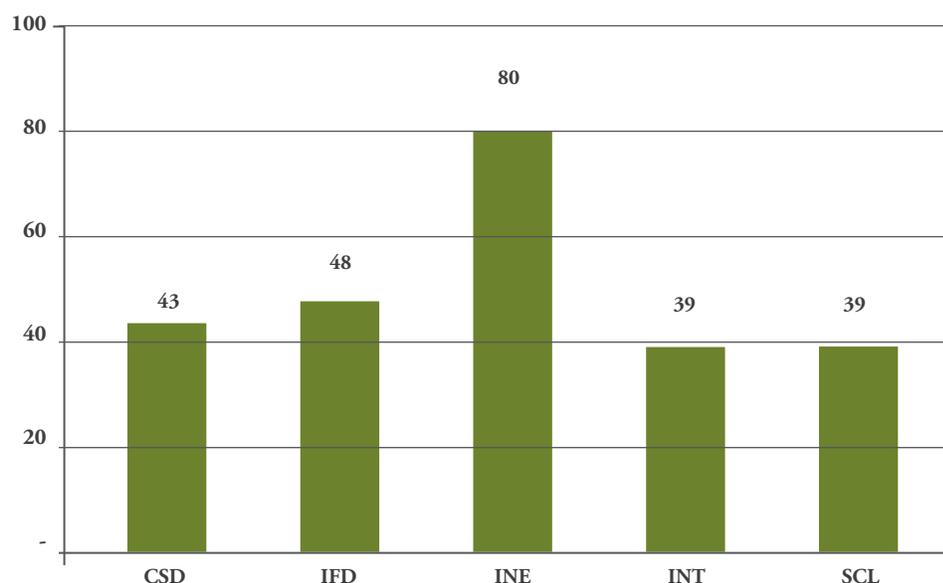


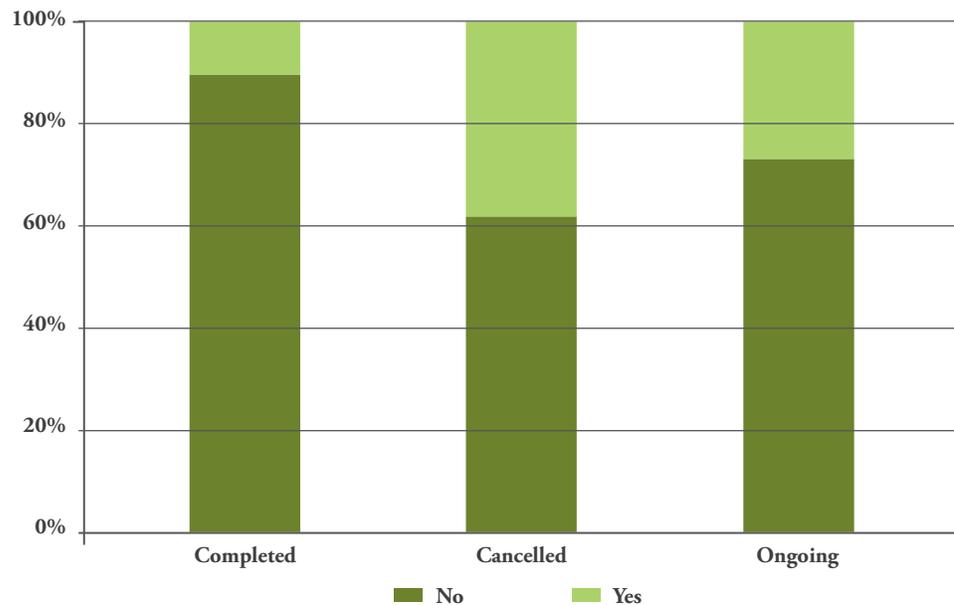
FIGURE 2.12

Percentage of IEs with SPD Support by Sector

Source: OVE. IE Database 2017.

FIGURE 2.13
Partnership to Design
and Implement an IE

Source: OVE. IE Database
2017.



In sum, despite progress over time, the Bank still needs to strengthen the IE origination process. SFDs provide heterogeneous guidance on knowledge gaps, explained in part by the sector’s level of maturity in analytical work. Many IEs continue to be Bank-driven, without full client buy-in. Sectors vary in expertise and experience, though those with less traditional engagement in IE have become significantly more active. SPD has helped to some extent by supporting sectors with less experience, at least in the design phase.

3. How much do IEs cost?

IE can be financed through four channels at the IDB.

- *Loan components.* Each project has a budget for “monitoring and evaluation”, and some use these resources to finance IEs.³²
- *Bank’s administrative budget.* IDB divisions can budget funds for financing IEs for completed operations that do not have prior allocation of resources, or they can complement loan funding with administrative budget.
- *TC funds.* TCs can finance IEs of IDB projects (complementing the amount already financed by the loan or funding the whole amount) or finance evaluations not directly related to IDB operations.³³
- *ESW resources.* Since 2009, IEs can be financed with ESW resources.

Most of the IEs considered in this analysis were financed by loans: 360 of the proposed evaluations (68%) were included in either an investment loan or a policy-based loan, and the remainder have been financed by TCs or grants.

OVE found large heterogeneity in IEs' budgets, which seems to be explained largely by the type of data used. OVE asked all staff responsible for IEs to provide budget data, which included all costs except Bank staff time. As Table 2.2 shows, 57% of IEs have budgets below US\$250,000, and 9% have budgets above US\$1 million. The main determinant of cost is not the methodology per se, but the type of data used for the evaluation (administrative or collected). The average cost of an IE based only on administrative data is US\$74,000, compared to US\$468,000 for IEs for which data are collected. According to interviews, important determinants of costs include the degree of dispersion of the beneficiaries and the country in which the evaluation is taking place. Countries with poor information systems are more likely to have data collected for their IEs.³⁴ It is important to note that the benefits of collecting data can go beyond the IE alone, especially in countries with limited information.

There is also heterogeneity among sectors in allocation of TC resources (Table 2.2). IFD and INE have used limited amounts of Special Program (OC) or Trust Fund resources to fund impact evaluations, while SCL has by far the largest share of these funds. Even inside SCL the distribution is unequal, with LMK and GDI using fewer resources than the other sectors.

TABLE 2.2 - DISTRIBUTION OF IE BY BUDGET AND TYPE OF DATA

Department	Division	Total IE. Method			Total Budget (US\$ Millions)				Avg. Cost of IE (US\$ Millions)
		RCT	Quasi-Exper.	Not Def.*	TC/Grant OC	TC/Grant T. Funds	Loan	Total	
CSD	CCS	1	2	-	-	1.0	1.0	1.0	0.65
	RND	1	17	2	1.7	4.5	18.8	20.5	0.47
	HUD	10	42	1	0.4	-	5.2	5.6	0.28
IFD	CMF	5	17	2	0.0	0.1	4.6	4.6	0.20
	CTI	5	33	9	-	0.3	9.4	9.4	0.21
	FMM	-	5	2	-	-	1.1	1.1	0.15
	ICS	21	39	12	1.9	0.8	7.7	9.6	0.14
INE	ENE	1	3	-	-	-	2.0	2.0	0.50
	TSP	3	11	-	0.2	-	9.9	10.1	0.72
	WSA	6	2	2	0.1	0.9	2.0	2.1	0.29
INT	TIN	1	21	1	0.1	-	0.6	0.7	0.03
SCL	EDU*	43	25	10	5.4	6.4	28.7	40.5	0.55
	GDI	11	5	4	1.1	3.1	1.0	2.1	0.26
	LMK	11	8	3	1.8	0.8	4.8	6.6	0.34
	SPH	49	40	27	5.3	7.3	54.6	59.9	0.58
	SCL	8	9	1	10.9	0.1	0.4	11.2	0.63
Total		176	279	76	28.8	25.2	151.8	187.0	387.6

Sources: OVE, IE Database 2017.

Note: *It refers to IE included in the loan documents or TC profiles, but the methodology was not completely defined in the M&E annex or the product matrix.

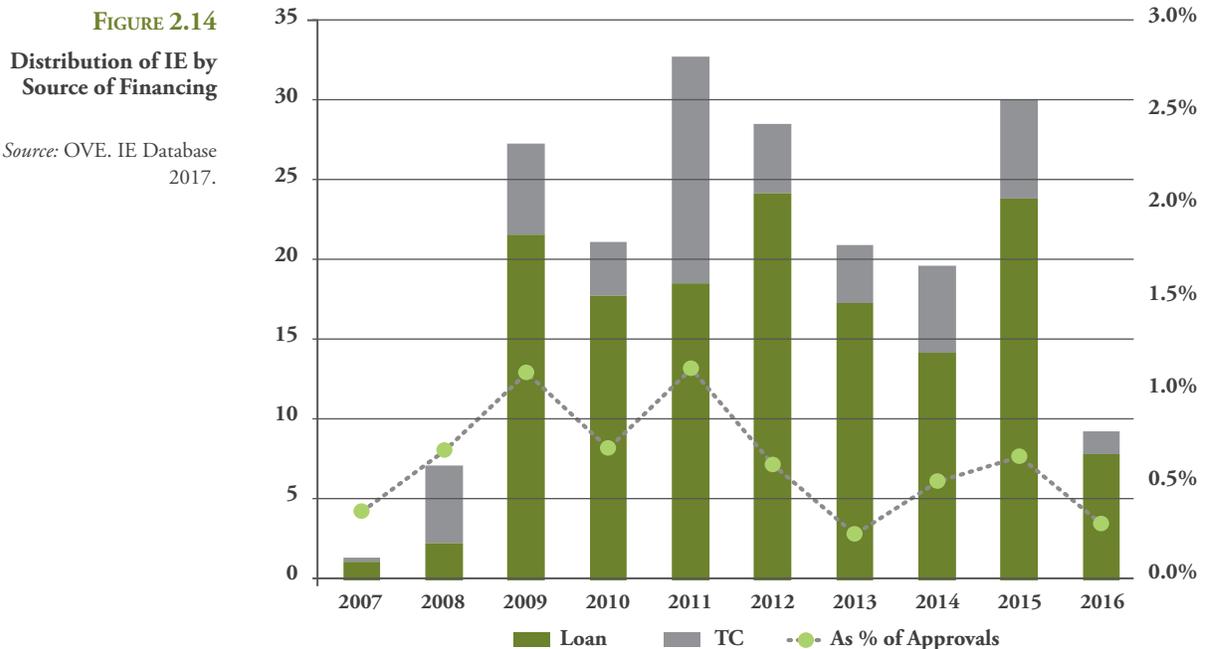
OVE compared the cost of IDB evaluations with what is suggested in the literature and observed in other organizations, and found that it falls in a reasonable range. Gertler et al. (2016) attempt to benchmark the cost of IEs by analyzing projects supported by the Strategic Impact Evaluation Fund administered by the World Bank. Like OVE, these authors find a broad range of costs (from \$130,000 to US\$2.78 million), according to the size and type of evaluation. Looking at IE cost in other organizations, OVE found that IDB’s budget distribution seems comparable with what DIME and J-PAL spend in their evaluations, but falls well below the MCC average.³⁵ DIME’s average is higher, as it relies heavily on experiments and data collection. MCC also finances more IE with data collection, and its evaluations cost about 3-5% of the cost of the project.

Overall, during the period of analysis, a total of US\$206 million has been budgeted to fund IE—US\$152 million from operations, loans and grants, and the remaining US\$54 million from TCs (Table 2.3). In operations that include at least one IE, the amount allocated to IE is on average less than 0.5% of the total amount approved for the operation (Figure 2.14). It is important to note that these estimates are based on available documentation and information provided by team leaders (rather than administrative records), and they do not include expenditures on cancelled IEs or on Bank staff time.

TABLE 2.3 - TOTAL BUDGET BY IE STATUS AND INSTRUMENT (US\$ MILLIONS)

IE	TC	Loan	Total
Completed	19.2	28.9	48.1
Cancelled	10.9	45.4	56.3
Ongoing	23.9	77.5	101.4
Total	54.0	151.8	205.8

Sources: OVE. IE Database 2017.



IDB-funded grants for IE total 60% of what countries spend through loans.³⁶ Non-reimbursable funds spent on IE total US\$77 million,³⁷ equal to approximately US\$0.61 for every US\$1 spent by countries (Figure 2.15). The distribution by country is heterogeneous, however, and there are no clear criteria guiding allocation. Paraguay, Chile, Dominican Republic, Jamaica, Mexico, Bahamas, and Honduras have received about average grant support. Other countries have received proportionally more grants per US\$1 of loans: Venezuela (US\$5.25), Costa Rica (US\$4.98), Ecuador (US\$1.71), Guatemala (US\$1.30), and Jamaica (US\$0.66). The other countries have received less than US\$0.61 per \$1. Many specialists indicated during interviews that requests for non-reimbursable funding were often motivated by a unit’s research agenda and did not necessarily coincide with priorities set by the country. Needs and interests of Bank sectors and individual specialists may also influence allocations.

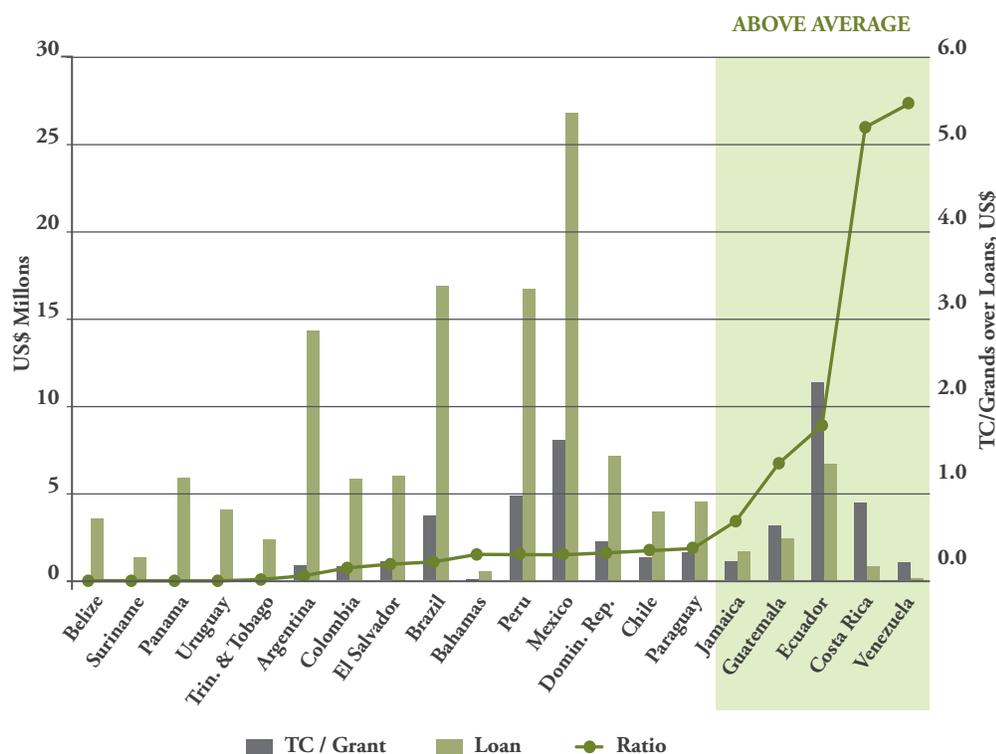


FIGURE 2.15
IE Budget by Country:
TCs, Loans, and Ratio

Source: OVE. IE Database 2017.

In sum, over the years the Bank and its clients have spent a lot of resources and accumulated a lot of experience with the financing of knowledge work, in particular impact evaluations. However, there does not appear to have been much strategic thinking behind the choice of evaluations to be conducted or the way TC resources have been allocated to finance them. The following section discusses some good practices in this area, based on the experience of several organizations.

4. What are good practices in financing impact evaluations and knowledge generation in general?

MDBs and other international development stakeholders are increasingly committed to producing and using IEs, and have used a variety of institutional arrangements for their financing. Drawing from the IDB's own experience, as well as the experiences of the World Bank, Organization for Economic Co-operation and Development, J-PAL, and Corporación Andina de Fomento (CAF), three key factors appear to promote effective and efficient financing of IEs: clear prioritization mechanisms, client participation, and cooperative networks.

A transparent and strategic prioritization process allows MDBs to target knowledge gaps in specific sectors, facilitates resource leveraging, and reduces the costs of implementing and disseminating knowledge by fostering long-term relationships. Donors are more willing to contribute funds when there are transparent allocation rules. Additionally, transparency in prioritization (as done by DIME) helps avoid donors' tendency to earmark resources. MDBs cannot produce all knowledge in-house and often subcontract specific products. Successful IE production models typically involve establishing strong ties with other knowledge producers like universities and specialized research centers, which bring important reputational gains for both parties and additional stakeholders in a cost-sharing structure. Moreover, long-term relationships with other knowledge providers can reduce the costs of prioritization by mainstreaming selection procedures, as done by J-PAL and IDB's Red de Centros.

It has been noted that successful models for IE financing involve active participation from knowledge end-users. Knowledge providers must have a strong link with clients, making them true participants from the IE design stage onwards. This link serves several purposes that contribute to best IE financing practices: it allows for the establishment of cost-sharing structures, builds capacity, and enhances dissemination efforts once the IE is completed. Cost-sharing is an important element for aligning incentives between different stakeholders; it creates ownership while allowing all parties to secure perceived savings. For example, DIME involves governments early in the process, and they pay 30% of the IE cost. Government counterparts receive instruction on IE methods and actively participate in the development and implementation of the IE. This generates ownership from government counterparts, creates natural checks and balances, and incentivizes the efficient use of funds. Furthermore, it is easier to disseminate an IE whose answer a client is willing to pay for.

Finally, given the nature of knowledge as a public good, establishing cooperative networks facilitates greater production of IEs. Knowledge exhibits increasing returns to scale. Cooperative networks enhance the diffusion and use of knowledge, which is necessary for increasing returns, while encouraging specialization to develop new knowledge. Strong cooperation schemes also make sense in terms of efficiency (e.g., avoiding duplication and competition). For instance, if IEs are provided with some level

of subsidy, cooperation ensures that the price is aligned among the different providers, eliminating arbitrage opportunities. Beyond increasing efficiency, cooperative networks allow leveraging resources both by efficiently using common resources and by more effectively attracting contributions from third parties. Through a cohesive ecosystem of several kinds of networks (formal and informal), with different temporality, varied objectives, and individuals in different disciplines, MDBs could maximize knowledge creation and increase the rate of return on their investments.

In addition to these general good practices, some organizations have developed guidelines and research protocols for defining when and how an impact evaluation should be pursued. For instance, the World Bank has determined that an impact evaluation would be useful when: i) it is testing an innovative intervention scheme, such as a pilot program; ii) the intervention is likely to be scaled up or replicated in a different setting; iii) the intervention is strategically relevant and will require a great deal of resources; iv) the intervention has not yet been tested; or v) the intervention results will influence key policy decisions.³⁸ Research protocols have been adopted by most organizations working with impact evaluations, and they have been shown to increase donors' interest in sponsoring evaluations.

Although IDB has in many ways been at the forefront of the promotion of IEs, it could benefit by incorporating some of these good practices. As noted above, the Bank lacks a transparent and strategic system to prioritize impact evaluations. The Bank could also benefit from promoting more client participation and ensuring, to the extent possible, their commitment. Finally, though partnerships have been established for training (with UC Berkley and Universidad de Los Andes, among others) and in a few cases for implementation (for example with J-PAL), more could be achieved through a more formal system of partnerships.

B. WHAT IS THE QUALITY OF COMPLETED EVALUATIONS AND THE DESIGNS OF ONGOING ONES?

As a first step in analyzing the quality of IDB's IEs, OVE looked at the number of publications arising from them. Of the 94 completed IEs proposed between 2006 and 2016, more than half are not documented in publications. Of the remainder, 15 were published in the IDB working paper series, 15 others were published in journals, and 9 more were published in both. This means that only 39 have been through a clear quality control process. According to interviews, the rigor and time required by the Bank's working paper series have made many specialists shy away from it.

OVE decided to look further into quality, recognizing there is no one unique way to assess the quality of analytical products. The large variation among sectors makes it difficult to compare across them, and over time. The prevailing consensus is that a quality assessment is best done through a peer review process in which different proposals are assessed by a group of professionals on different dimensions, such as adequacy of data and method and contribution to the literature.³⁹

To have a more comprehensive review of the quality of the IE produced by IDB, OVE hired CLEAR-LAC⁴⁰ and developed a quality framework that summarizes the key elements of the identified IEs.⁴¹ CLEAR-LAC peer-reviewed a sample of 86 completed evaluations⁴² and 59 proposals of ongoing IEs,⁴³ providing for each an overall assessment of its quality and its potential to be published in a journal or as a working paper. (Annex III provides more details on the sample selected, the method used and the findings from CLEAR.) The quality assessment was based on four evaluation dimensions:

- *Relevance*: the quality of the literature review, the relevance of the questions addressed by the evaluation, and the quality of the research protocol.
- *Methodology*: the selection, rigor, and shortcomings of the analytical method, the timeliness of the post-treatment period and follow-up, and the existence of power calculations.
- *Data*: sample size, unit of analysis, and type and quality of data.
- *Results*: the analysis result and implementation problems such as noncompliance, attrition, and spillovers, as well as their possible effects on the reliability of the evaluation.

CLEAR's assessment found that the quality of 55% of completed IEs is overall satisfactory or partly satisfactory, while the remaining 45% are unsatisfactory or partly unsatisfactory. Many factors affect this overall assessment. Only 17% of completed IEs present an exhaustive literature review and include a good-quality research protocol. Furthermore, only 42% of completed IEs exhibit rigorous methodological approaches, while 23% present serious methodological shortcomings. In this sense, it is interesting to highlight that not all methods are applied with equal rigor. RCTs are the most popular single methodological approach, used in 35% of the IEs. The rest use other quasi-experimental methods (56%) or multiple methods (9%). While 60% of RCTs were found to be methodologically rigorous, only 30% of quasi-experimental IEs showed the same rigor. Moreover, RCTs result in better data quality than other methods: 70% of RCTs, and only 55% of all other methods, have detailed descriptions on data collection and variable construction. Although 72% of the completed IEs have a satisfactory or partly satisfactory sample size, 78% do not have or do not discuss power calculations; this is pervasive across all divisions. Finally, the assessment shows that implementation issues like attrition and spillover effects affected the reliability of at least 34% of the IEs.

Quality differed significantly by sector (Figure 2.16). As was mentioned in the portfolio analysis, most IEs in divisions that did not traditionally work with IEs have not yet been completed. Three sectors (FMM, TSP, and CMF) completed three or fewer IEs in the period. The difference in their overall quality assessments reflects the complexities in the design and implementation of IEs in these sectors.

This highlights the importance of extracting lessons from failed IEs to inform future IE production, particularly in sectors with less experience. On the other end of the spectrum, three divisions (LMK, SPH, and EDU) produced 14 or more IEs during this period. These divisions show a more balanced distribution in terms of quality, with LMK being the top performer by having only one IE of unsatisfactory quality. Mid-producers (GDI, RND, CTI, and ICS) have, on average, the lowest percentage of IEs with unsatisfactory quality and the highest percentage with satisfactory and partly satisfactory quality. This could be associated with better prioritization mechanisms, given the constraints in their capacity to generate IEs compared to big IE producers. OVE interviewed all division chiefs to gain a better understanding of the IE production processes by division. For GDI and RND—divisions with the highest percentages of partially satisfactory and satisfactory IEs—OVE’s findings corroborate the prioritization hypothesis.

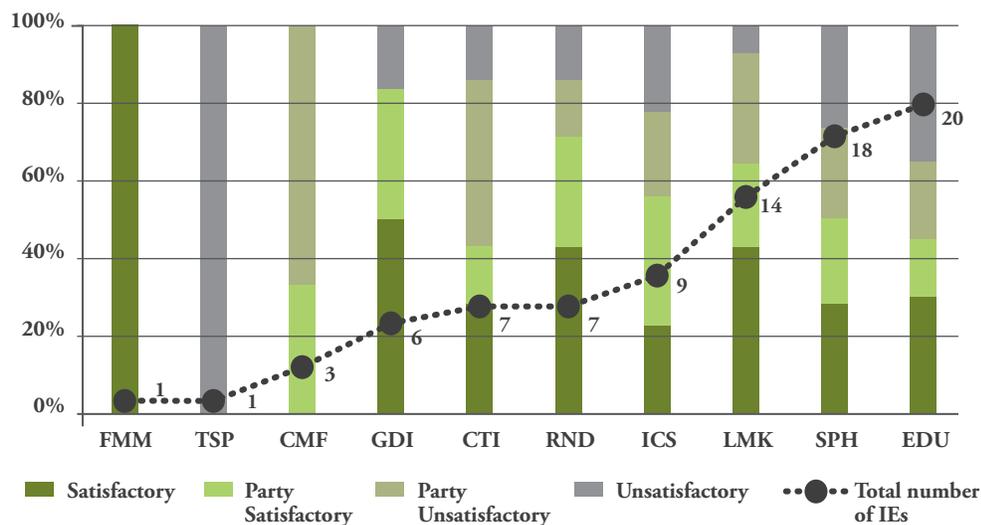


FIGURE 2.16
Quality Assessment of IEs by Division

Source: OVE. IE Database 2017.
Note: 2014 and 2015 were not included in this graph because of the small number of IEs proposed these years that are already completed.

While quality trends are difficult to disentangle over time, it is fair to say that the quality of IEs has increased. Figure 2.17 shows the effects of the DEM on quality. Before 2009 certain sectors, such as SPH and EDU, were farther along the learning curve on how to design and implement high-quality IEs, and they were the only sectors that included IEs in their loan and TC operations. However, the introduction of the DEM in 2009 strongly incentivized the production of IEs and made “new” sectors enter the IE market. This translated into a decrease in the average quality of IEs proposed between 2009 and 2010, mainly driven by less experienced sectors. After 2010 there seems to be an upward trend in quality, an indicator that overall sector IE experience is increasing. Furthermore, judging from a representative sample of proposals,⁴⁴ the quality of IE proposals has improved over time (Figure 2.18), with better literature reviews, stronger research protocols, and better analytical methods. An improvement in the quality of proposals is a step in the right direction, since it is the best proxy for the overall quality of the end-product.

FIGURE 2.17
Completed IEs: Quality Assessment by Year

Source: OVE, IE Database 2017.

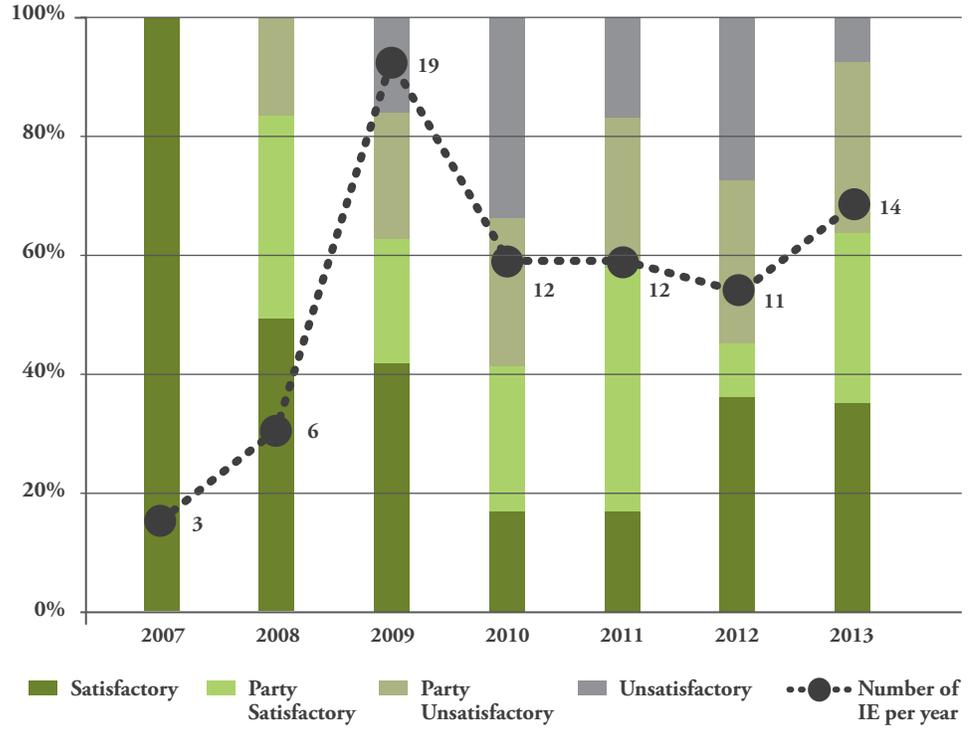
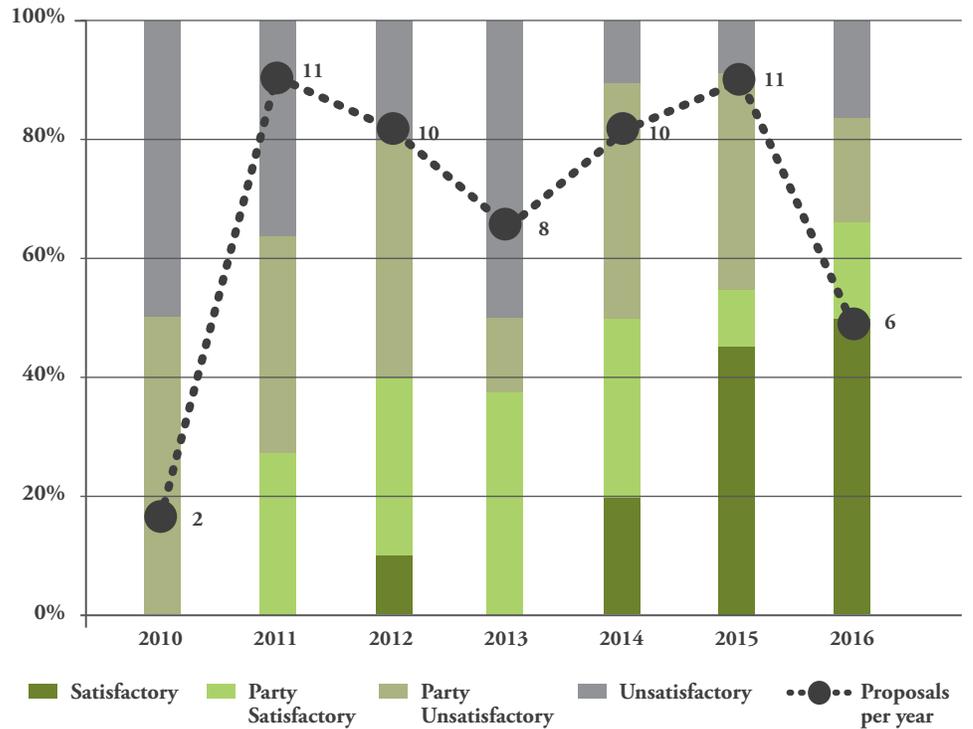


FIGURE 2.18
IE Proposals: Quality Assessment by Year

Source: OVE, IE Database 2017.



This increase in quality might be explained by several factors, including some effect of “learning by doing” and a strong effort in IE training. SPD has put together an evaluation hub with an impressive amount of material on impact evaluation (for all the steps from design to dissemination), and has organized extensive training for IDB staff and clients. In 2015 and 2016 alone, SPD/SDV taught 40 courses and workshops and organized almost 50 more, with more than 2000 attendees.⁴⁵



3

One of the main motivations behind the push for IEs is the improvement of IDB's capacity to promote effective development in the region.

© IDB

3 Use and Influence of IDB's Evaluations

This chapter focuses on the use and influence of the IDB IEs. It is important to keep in mind that this type of analysis is generally very complex because there is no unique way to measure use and influence. As a form of knowledge, IEs could be used to (i) inform IDB work and strategies; (ii) inform policymakers in the region, supporting sound policy decisions; and (iii) help close knowledge gaps in general, reaching a broader audience. This chapter analyzes each of these uses, as well as a potential positive externality of IEs—their use as a tool to support project design and implementation.

A. USE AND INFLUENCE OF IMPACT EVALUATIONS AS KNOWLEDGE

1. To what extent are IEs used internally in IDB?

One of the main motivations behind the push for IEs is the improvement of IDB's capacity to promote effective development in the region. In this sense, it is important that the IEs produced be used internally. OVE attempted to understand how staff members are using evaluations, and how they are integrating them with the project cycle and the programming process. This assessment is challenging, as IEs serve different purposes and audiences simultaneously, and not very many have yet been produced. OVE conducted semi-structured interviews with staff members across VPC and VPS to analyze the use of evaluative results by staff involved in IDB operations. These interviews brought several insights on how IEs are used in practice, and what their relative importance is in comparison with other analytical products.

An analysis of the SFDs indicates that the IDB does not systematically use IEs to inform and help define broad approaches at the sector level. The use of the evidence produced through IEs can improve the theory of change behind IDB operations, promote organizational learning, enrich the stakeholder engagement strategy, and inform the design of ongoing monitoring. Though SFDs are not required to specifically account for and guide IEs, they should include any information deemed relevant to improving the quality of the Bank's operational and analytical work at the sector level. Furthermore, they are updated periodically (every three years) to promote continued learning, which makes them the logical instrument through which emerging IE evidence can feed into decision-making. However, 40% of current SFDs do not meaningfully include IEs when citing international evidence at the sector level.⁴⁶

Interviews revealed that there are two stages in a loan cycle in which IEs are more likely to be used: first, during the programming process, when an agenda of loans is discussed with the country; and second, during project design, when specific interventions are planned in detail. At both points the experience from previous operations and the findings from evidence-based literature are useful.

During the country programming process, results of previously implemented programs appear to be used alongside other evidence-based documents. Country strategies provide the basis for future programming, and the ideal strategy process lasts about two years, starting one year before the change of office in any given country, and ending one year after. In general, discussions in the first year revolve around the role the Bank could play in the country's different sectors, and the final agenda of loans is set in the second year. Several analytical instruments (growth diagnostic report, country program evaluation, and policy and sector/thematic notes) help diagnose the needs of the country and provide a technical background to the negotiation on specific issues. OVE looked at the sources of evidence used in country strategies and sector/thematic notes and found that IEs are not frequently cited. Some sectors (SCL, INT) use IEs more frequently, alongside evidence-based literature from academic journals, while others (INE, ICF) use mostly other analytical inputs generated on a project-by-project basis.

Despite the limited evidence of IE use in country strategies, technical notes, and SFDs, interviews have suggested an important use of IEs in the country dialogue: to avoid programs with limited effectiveness. IDB support is mostly demand-driven, and country-specific interests, sometimes politically-driven, are highly relevant. In some cases the presentation of empirical evidence on effectiveness can help guide governments toward different and better interventions. The "One Laptop per Child" program is a case in point. The program—which represented about one-quarter of the overall expenditure on IE at IDB—had been implemented in a few countries when the results of the IE showed no impact on children's achievements. As a result, the program was cancelled or not approved in at least three countries (Peru, Chile, and Mexico).

The use of evidence from IDB IE during project design is still very heterogeneous. SCL and INT actively use IE results more actively than other sector units for several reasons. First, only a relatively small number of completed IEs address issues that are relevant for other sectors. Second, many of the evaluations that do exist are of moderate quality and do not necessarily concentrate on the most important questions for project design. Third, there is some inertia in how the Bank works, and similar project designs are often applied across countries regardless of the lack of evidence of their effectiveness or, even more worrisome, sometimes with evidence of lack of effectiveness.

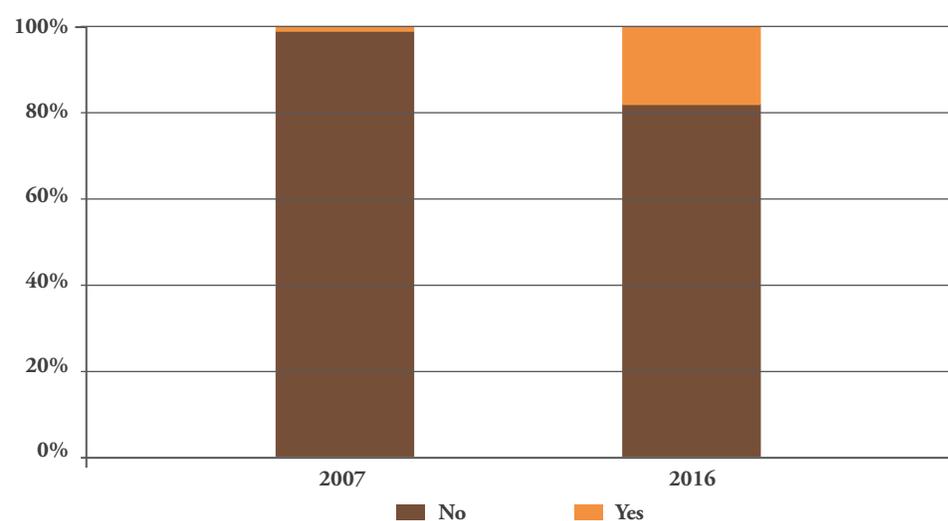


FIGURE 3.1
Operations whose
Documentation Cites IE
Results to Support their
Design

Source: OVE. Analysis of
Publications.

In addition to conducting interviews, OVE analyzed the references to IEs in loan documents produced by the Bank, confirming the heterogeneity in their use across sectors. OVE reviewed all loan documents approved in 2007 (before the introduction of the DEF) and 2016 (the most recent available) for citations of IEs in the description of the intervention design. SCL projects are more likely to rely on evidence from IEs, but overall the analysis found that all sectors increased their use of IE in project design (Figure 3.1).⁴⁷

2. How are IDB IEs used by clients?

IDB IEs could have a key role in policymaking among clients, but it is hard to assess it. They might facilitate real-time feedback to policymakers and help to determine the most effective design for future programs.⁴⁸ IE can also be used to provide external validity to projects. They can provide an objective view, and they provide rigorous evidence that can help policymakers decide, for example, whether to expand, modify, or eliminate programs. OVE believes that the best approach to assess the role of IEs in policymaking would be to measure the context of changes in public policy over an extended period. However, such an approach would have

two limitations: (i) it is not possible to isolate the influence of IDB IEs from other sources of knowledge; and (ii) IDB production of IEs is still too recent. The main longer-term process that we can observe in practice is PROGRESA/Oportunidades (Box 3.1).

Box 3.1 - It is not one IE that makes the change: The Case of PROGRESA/Oportunidades

PROGRESA/Oportunidades is considered an example of results-based policy. The program was rigorously evaluated, and the model has been applied in several other countries. The evaluation of the program served to raise the cost of discontinuing the program when a government from the opposition took power. It also served to foster legitimacy in the intervention. These are roles that go beyond knowledge.

However, evidence from sources other than IEs played an equally relevant role in the formulation of the program. Decisions on structuring payments, conditioning, and making payments to mothers—among many others—were informed by evidence, gathered over decades, on the dynamics of household allocation of resources. This means that while evidence did play a role in PROGRESA, the mechanism was based on an accumulated body of work in development, and not necessarily on specific IEs. What is not as clear is the extent to which the IEs led to changes in the program. Behrman (2010) shows that there are mixed opinions, but substantial skepticism regarding which of the changes were truly influenced by evaluation.

By all accounts the PROGRESA IEs have had a much more relevant role as a public good than they have in terms of the program's operation. A recent review of the program, including extensive interviews with program designers and evaluators, shows a unanimous opinion that the IE program was key in replicating the model in other countries and in promoting its continuity in Mexico. The evaluation strategy helped to make conditional cash transfer programs (CCTs) more palatable and attractive, and many later CCTs were heavily influenced by earlier ones, especially PROGRESA/Oportunidades. It should be noted that the influence stemmed not so much from the content of the evaluations as from the fact that they were done and were largely regarded as independent and credible.

OVE interviewed a sample of IDB clients who have been involved in projects with IEs, and most said that the main benefit of IEs is the possibility to adjust the functioning of ongoing policy. OVE selected a stratified sample by department of 50 IEs (completed and ongoing) to follow up with the Bank's counterpart at the government. However, OVE able to reach at least one person in the government to be interviewed for only for half of this sample. The result was a sample of 25 IEs.⁴⁹ A total of 52% of the interviewees saw IE primarily as a tool to adjust policy, while 20% identified it as an instrument to support and validate existing policies. Interestingly, some stressed the importance of IEs as an accountability tool to justify borrowing from IDB.

OVE found a wide range of views about the usefulness of the IEs. About 56% of interviewees stated that the IE influenced or was expected to influence policymaking, while 40% said they had used it or were expecting to use it in the discussion of other policies. Among interviewees whose IE was ongoing (as opposed to finished), the use or expectation of use in current policy rose to 83%. Only one interviewee found no value in the IE and considered it a cost, and a second believed that IEs in general are not carried out with the quality needed to isolate effects and inform policy. For those that admitted little to no use of the IE to adjust policy, political issues and administration change were recurring factors. In addition, issues with timing of the IEs have been identified as a barrier to use. (Box 3.2 provides a broader discussion on the limitations to use.)

Box 3.2 - Impediments to the Use of Impact Evaluations in Policymaking

Despite IEs' potential usefulness of impact evaluations in policymaking, there are many reasons for them to be underused.

- Competition with other sources of evidence (Ravallion, 2008). IEs are often written with high academic rigor and focus on method rather than on policy implications, so that it is rather difficult for the non-academic public to use them. This problem can be overcome, to some extent, through strategic dissemination of completed IEs. INT provides a good example: after several evaluations are completed, the sector releases books focused entirely on disseminating lessons learned (excluding the technical discussions).
- Different focus from policymaking (Ravallion, 2008). IEs often focus on the question of what works, while policymakers are more concerned with cost-effectiveness. To address this issue, an effort should be made to carry out better IEs that go beyond pure impact, treating issues of effectiveness. This paper shows that IDB is moving in this direction, responding more to the questions that policymakers are interested in, rather than just using IEs as an accountability tool for its projects.
- Timing issues (Sutcliffe and Court, 2005). Policymakers often need to advise ministers on pressing issues. Unfortunately, both the production and subsequent publication of scientific knowledge take time—a gap that may make policymakers substitute less rigorous methods, assumptions, or personal or institutional values without waiting for solid empirical evidence. This trade-off between timely policy findings and rigorous evidence should be kept in mind (especially by development banks); by the time the evidence produced is finally available, it may no longer be relevant.

Interviewees were more likely to perceive an IE as being useful when the government had been directly involved in the design of the evaluation. As Weyrauch and Diaz Langou (2013) suggest, the communication among relevant actors is indeed key to



determining how research reaches the policy field. The Bank seems to be internalizing this fact and increasing its communication with the client to better fit the IE into the client's knowledge needs. For instance, all the interviewees corresponding to ongoing evaluations stated that the IE fit their evidence needs.

Capacity building and increased interest in integrating IEs into governments' M&E systems are two positive externalities identified by interviewees. More than half stated that the IE's data requirements led to improvements in their M&E system or in their reporting of activities. Furthermore, two-thirds of the interviewees agreed that after the experience with the IDB IE, they are better equipped to conduct IEs in the future, partly because of the training provided by the Bank.⁵⁰ Even those that did not have training feel more prepared to hire consultants to do the work.

3. What is the broader reach of IDB IEs?

Impact evaluations can be influential only if they are used; but for results to be used, they need to be known. Of the Bank's completed IEs, 55 (58%) were not published in journals or IDB working paper series.⁵¹ Many of them could not even be found in IDBDOCS, and OVE had to request them directly from the team leader. This shows that IDB's IE results are not easily accessible to the public.

OVE also did a citation analysis, comparing the frequency and patterns of citations of articles and books produced by the Bank with those of other MDBs.⁵² Between 2006 and 2016, IDB-affiliated authors published 177 articles citing IEs;⁵³ these IEs were cited a total of 2603 times, averaging 16 cites per paper and 237 cites per year of publication. The years in which the most IEs were published were 2012 (23) and 2015 (27), and the publication years with the greatest number of citations were 2012 (515), 2013 (377), and 2011 (312). However, as Figure 3.2 shows, the average number of citations per paper for IEs published in 2015 (3) is far lower than for those published in 2012 (26), possibly because the IEs are more recent and have not yet appeared in follow-on research. The citations have mostly appeared in IDB technical notes, and a few authors are responsible for most of them. The 14 most published authors account for almost 90% of the citations in this period.

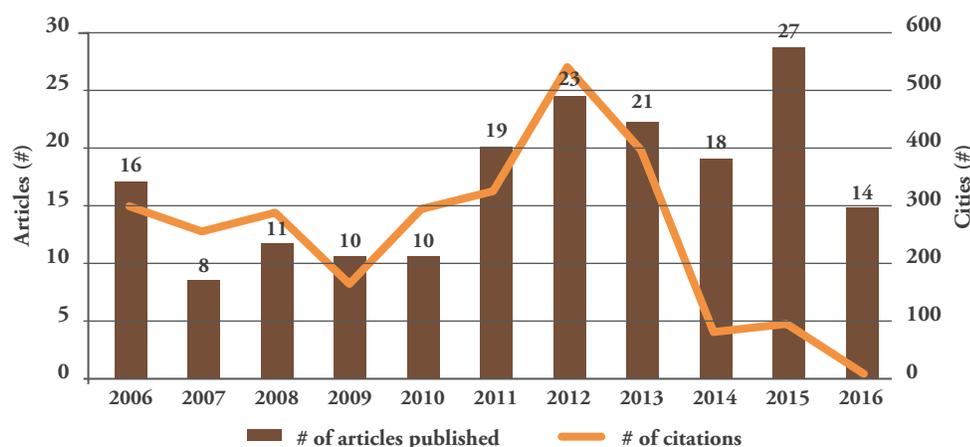


FIGURE 3.2
Publication and Citation of
IDB IEs by Published Year

Source: OVE, based on Publish or Perish citation results.

To put this analysis in context, OVE looked at the citation of general publications of IDB vis-à-vis those of other organizations, and found that IDB's performance is about average relative to its comparators. The average number of citations per publication for the IDB is 7.9 for all publications between 1996 and 2010. For the same period of analysis, the average number is lower for the ADB (4.2), AfDB (3.4), EIB (6.8), and IFAD (3.9), but higher for the World Bank (13.5), IMF (8.9), UNDP (18.6), and EBRD (11.2).⁵⁴

B. CAN IE HELP IN DELIVERING PROJECTS?

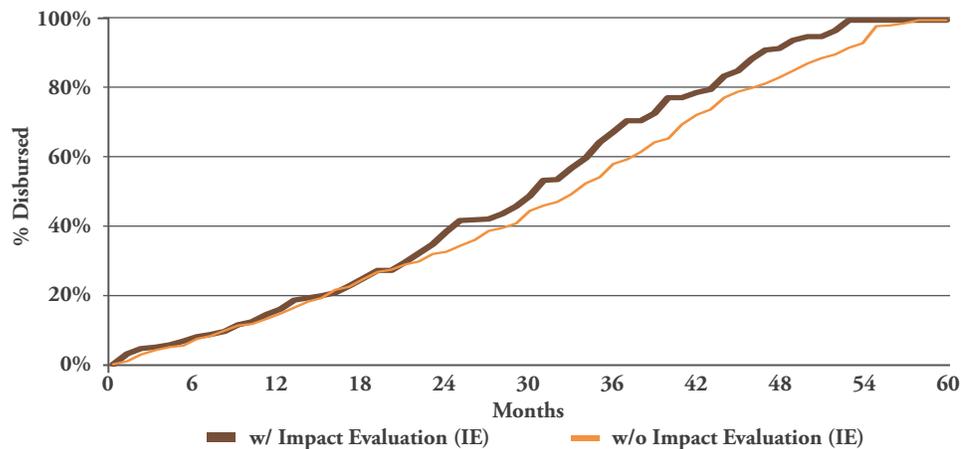
Most of the effort to increase the production of IEs in IDB was motivated by the idea of strengthening knowledge and ensuring accountability. However, there is evidence that, when properly designed and justified, IE can also help project implementation. Legovini, DiMaro, and Piza (2015) use data on project implementation at the World Bank to show that projects that have IE disburse faster. The authors list several channels through which an IE could be helping project implementation—for example, through better planning and evidence-based design and through better

implementation capacity brought about by the work of the research team and field staff. They also argue that the process of preparing the IE increases the amount of data for policy decisions, thus helping the project and generating “observer effects and motivation,” as knowledge about the IE can generate expectations.

OVE found evidence that IDB projects with IEs that were approved between 2009 and 2016 disbursed faster than those without IEs.⁵⁵ There is a clear difference between the disbursement curves of projects with and without IE (Figure 3.3). While they seem to be almost the same during the first two years of implementation, projects with IE speed up disbursements after that and are completed earlier. Projects with and without IE have similar planned durations on average, but those without IE have more delays (on average three months more) than those with IE. Annex VII includes the econometric analysis and shows that the difference is significant and robust even when controlling for many characteristics of the project, such as amount of the loan, sector, country, number of team leader switches, year of approval, and even project logic DEM score (as a proxy for project quality).

FIGURE 3.3
Disbursements of
Investment Loans
with and without IE

Source: OVE. IE
 Database 2017.



The exact channels through which IEs could help project delivery are not clear. OVE attempted to shed some light on this by comparing the disbursement rates of projects with IEs that were completed or are ongoing and likely to be completed with projects with IE that have been cancelled or are on hold. Figure 3.4 is analogous to Figure 3.3, but it separates the projects with IEs into two cases: (i) projects with completed or ongoing IEs (“successful” cases), and (ii) projects with cancelled or on-hold IEs (“unsuccessful” cases). There is a minor difference in the disbursement curves of projects without IE and projects with cancelled IEs, but the previous result holds for projects with completed or ongoing IEs. This supports the view that IEs play a role in project implementation, likely due to the support and training provided by researchers and field staff following the project. Also, the specific support from governments that IE requires could potentially be one of the drivers of the results presented here.

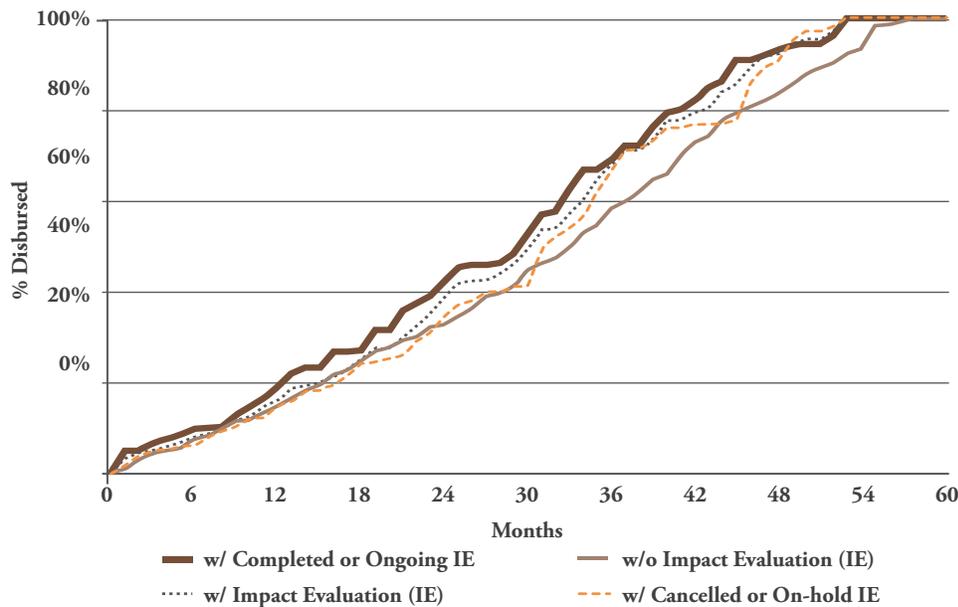


FIGURE 3.4
Disbursements of
Investment Loans
with and without IE

Source: OVE, IE
Database 2017.

The potential better performance of a loan with an IE has financial implications and could offset up to half the cost of the IE.⁵⁶ OVE estimated these savings using the average amount of the loans with and without IE and the current credit fees paid for undisbursed balances. The average savings from faster disbursement are around US\$40,000 a year. Taking an average project cycle of five years, the total amount saved by the clients would vary from 25% to 50% of the average cost of an IE.⁵⁷ On the basis of this analysis OVE concludes that IE, when well justified, besides complementing knowledge in some areas, may have a positive externality in project implementation and in reduction of financial costs.



The Bank has made progress towards better-quality evaluations, but there is still room for improvement.

© IDB

4 Conclusions and Recommendations

This document takes stock of the effort to promote IE practice in the IDB. The first finding is the lack of systematized information regarding the IEs. After more than 10 years investing in IEs, the Bank still does not have a centralized system to track all IEs in progress and their cost, nor does it have a basic registry of evaluation questions, compliance with ethical standards, and research protocols. SPD has developed a system to track IEs linked to loans, and is currently working on a system for IEs included in TCs. However, IEs financed by ESW funding cannot be easily tracked, nor can the exact cost of the evaluations. Furthermore, there is not yet a single repository for reports and datasets.

Bank incentives in recent years have led staff to over-propose IEs. The DEM scoring system influenced many teams to propose evaluations that did not always meet the necessary conditions for implementation. Two consequences should be highlighted. First, many evaluations have been cancelled, limiting information about the results of projects that have poor M&E arrangements. OVE could not assess the monetary cost of initiating these failed IEs. Second, the strong incentives in the DEM seem to have promoted a cultural shift in the Bank: in general, divisions are putting more emphasis on development effectiveness and how to assess it—if not through rigorous IEs, at least with a counterfactual mindset. In addition, in the divisions that have been working longer with IE, a clear learning process can be identified. There now seems to be more strategic thinking behind the evaluations these divisions are proposing. Others are in an earlier learning stage, still attempting to identify where to use IEs. However, some IE continues to be proposed in an ad-hoc fashion following the interest of the specialist in charge of the project. This suggests that the Bank still needs to be more strategic in defining the IEs to be conducted.

The range of costs of IEs is broad, but costs are broadly comparable to those of other organizations. IDB's clients bear most of the cost of IEs. The main driver of IE costs is data collection, and in some cases such data may have other uses for the client. The nature of all knowledge products as public goods may lead to underinvestment, however. When there is no need to collect data, costs are much lower, and the main issue is how to find an appropriate combination of incentives and services to facilitate their funding and delivery.

The Bank has made progress towards better-quality evaluations, but there is still room for improvement. The quality of the IEs was judged following international standards, considering the relevance of the evaluation question, the appropriateness of the data used, the rigorosity of the method, and the robustness of the analysis. About 55% of the completed evaluations were considered to be of satisfactory or partially satisfactory quality. The analysis shows that after the introduction of the DEM the quality decreased, as less traditional sectors started to do IEs, but it continued to increase afterwards. Similarly, the quality of the proposals shows an upward trend.

Despite the strong push to produce IEs, they are not widely used in IDB documents, though they can be an important source of information for discontinuing ineffective programs. Most of the 25 clients interviewed for this report indicated that the inclusion of IE in the operations was useful. Many suggested that these evaluations supported the continuity of the program evaluated and facilitated its accountability. A few suggested that IEs are imposed by the Bank and are little used. OVE found little evidence of IE use in country strategies, technical notes, and projects, as measured by direct citation, though interviews suggest that the IEs have a role in providing evidence to discontinue ineffective interventions or programs.

Finally, the analysis provides some evidence that IEs could positively affect project implementation. Though the causal mechanism is not clear, this could result from better project design (to allow for an IE to be implemented) or stronger support from the evaluation team in the field. IDB projects with IEs disburse faster and finish earlier than those without IE, potentially offsetting up to 50% of the evaluation cost.

Reflecting the findings of this evaluation, OVE has the following recommendations to improve the production and use of IEs in IDB:

1. Be more strategic in the selection of IEs by undertaking or supporting IEs only if they have client commitment (or a clear strategy to engage the client to develop such commitment), a well-identified knowledge gap, and a feasible timeframe. This evaluation has shown the importance of having the client on board when conducting an IE. Client commitment also helps to ensure the use of IE results at completion.

2. Ensure that SFDs identify knowledge gaps to help guide the Bank's IE work. SFDs have done this to varying extents to date, and OVE recommends that they do so more consistently and thoroughly going forward.
3. Revise the DEM to reduce the incentives to over-propose IEs. Every project should count on a solid monitoring and evaluation system, but not all of them need to have an impact evaluation. The DEM has been a useful tool to increase projects' evaluability and to promote more understanding of the importance of doing more rigorous evaluations. It has also increased the incentives to propose IEs, leading to proposals that were not well-thought out or warranted.
4. Develop a transparent funding mechanism, aligning the interests of clients and the Bank. When an IE is deemed relevant and feasible for a loan, it should be incorporated into the M&E system funded by that loan to the extent that it will be informative for the country and for the preparation of the Project Completion Report (PCR). If there are longer-term impacts that can be measured only after the project closes, post-project evaluation work should be funded by Bank-managed resources. In this way clients who are committed to IEs and put in place proper M&E arrangements could benefit from IDB resources to complete the evaluation after the project closes. At the same time, the Bank would commit its own-managed resources only to those longer-term evaluations where there was client commitment, good early management (with proper baselines and follow-up), and an expectation of valuable information on longer-term program results.
5. Strengthen systems for quality control. The Bank already has a number of mechanisms in place for ex-ante quality control. However, as shown in this evaluation, not all IEs are completed with high quality. To avoid the reputational risk of publishing and disseminating low quality work, the Bank needs to develop a system for quality control beyond the IEs published as Working Papers or Technical Notes.
6. Strengthen and move towards centralization of the Bank's systems to track all impact evaluations being conducted—from design to completion—and their costs, and to make high quality IEs and their documentation, the data used (when possible), and their lessons available to the public. SPD and KNL have made progress in these directions, as noted in this evaluation, and the system should be completed.
7. Develop a more formal mechanism to promote partnerships to increase the quality, relevance, and visibility of IDB work, while potentially reducing costs. This evaluation found little partnership in the actual production of IEs. Based on what has been observed as good practice in IDB and other organizations, the Bank could usefully increase its partnerships with academia as well as its collaboration with other organizations in IE production, dissemination, and information-sharing. SPD has been promoting this, and the effort should be strengthened.

- ¹ Cameron, Mishra, and Brown (2016) constructed a massive database of IE published between 1981 and 2012 to take stock of “how much we have learnt.” They showed that, even though IE was initially used mostly in the medical literature, since the 2000s it has been increasingly used to assess policies in many different countries—with some concentration in Latin America and in Asia—and using a variety of methods beyond experiments.
- ² Center for Global Development (2006b).
- ³ “What Have We Learned? Improving Development Policy Through Impact Evaluation.” Seminar <https://www.povertyactionlab.org/event/what-have-we-learned-improving-development-policy-through-impact-evaluation>.
- ⁴ See Cunha, Heckman, and Schennach (2010) and Bedoya et al. (2017).
- ⁵ An IDB loan was used to fund Mexico’s PROGRESA. Levy (2004) and Levy and Rodriguez (2005) describe how this program was designed, implemented, and evaluated. Mexican authorities created the conditional cash transfer program using evidence from previous decades on the dynamics of household allocation of resources. However, they did not know the specific effects of the program, and an IE was designed for this purpose. The evidence collected with the evaluation was key in replicating the model in other countries, and in promoting its continuity in Mexico, mainly through IDB support.
- ⁶ “The Inter-American Development Bank has played and continues to play an active role in the establishment of these [targeted human development] programs and in ensuring that they incorporate sound impact evaluation systems” (Legovini and Regalia, 2001).
- ⁷ “Very few projects defined meaningful indicators of performance, while very few had either baseline data or explicit targets and milestones. Where such empirical indicators and targets did exist, they were applied to project outputs only, not to the outcomes anticipated to be realized in the country once the project had produced its outputs” (OVE, 2002).
- ⁸ OVE’s IEs were typically conducted ex-post given its independence from day-to-day operations. For a discussion of the challenges of doing ex-post impact evaluation and how OVE addressed them, see Ruprah (2008).
- ⁹ Between 2006 and 2012 the share of new hires with master’s degrees increased from 54% to 59%, and the share with doctoral degrees increased from 13% to 23% (OVE, 2014).
- ¹⁰ For more details, see IDB document GA-232.
- ¹¹ The DEF is currently a Bank Operational Policy.
- ¹² See AB-3008, paragraph 4.23.
- ¹³ The challenges of financing IEs in the public and private sector are very different. At the time of this evaluation, IIC had one completed evaluation and a small number in pipeline. According to OVE’s evaluation of the MIF (MIF/RE-2), MIF has taken steps to strengthen its project preparation, monitoring, and evaluation, and has invested in an IE program. However, it fell short of its target of evaluating 25 projects a year.
- ¹⁴ IEs can be produced (i) as a pure research product, (ii) during project preparation and implementation, (iii) as part of stand-alone sector work, or (iv) ex-post as part of the Bank’s normal project evaluation cycle.
- ¹⁵ This does not necessarily mean that no evaluation was done. In most projects that were not cancelled, the IE was replaced by another type of analysis.
- ¹⁶ Of the operations with IEs, 105 are scheduled to be completed in 2017, with decreasing numbers in the following years—98, 58, 28, 36, and 3. However, the completion of the IEs usually occurs some time after project completion.
- ¹⁷ Despite the generally high response to OVE interview requests, there was a higher non-response rate among team leaders of cancelled IEs.
- ¹⁸ Team leaders were asked about the reason for proposing each IE—interest from the Bank or the client, the DEM score, or others. If the response was the DEM, the probability of cancellation was 10 percentage points higher.

- ¹⁹ In 18% of RCTs the identification strategy was changed to another method.
- ²⁰ This evolution can also be seen outside IDB, and more effort has been put into sectors that traditionally did not rely on IE (Cameron, Mishra, and Brown, 2016).
- ²¹ Other departments have also made substantial use of ESW resources to finance their IE, but most of them have been identified through TC operations included in the analysis.
- ²² During recent decades, methods to collect data—census and surveys—have focused on social indicators, facilitating access to quality data that can be used to assess interventions in these sectors.
- ²³ For detailed information see Annex II, table B.1.
- ²⁴ Many of the cancellations were replaced by ex-post economic analysis. The lack of consistency in the responses to OVE’s interviews made it impossible to identify every case.
- ²⁵ This is particularly critical in sectors and countries with limited administrative records or national surveys. The information collected through the IE represents the only source of information to evaluate the impact of the intervention. An example of this is the conditional cash transfer programs in Central America.
- ²⁶ Gertler et al. (2016), Kaplan (2016), and Karlan and Appel (2016).
- ²⁷ This finding comes from interviews with a sample of IDB clients involved with IDB’s IEs.
- ²⁸ The percentage of consultants is higher in INE and INT, where consultants do more than 85% of the evaluations. In these cases counterparts’ participation in design and implementation is still very limited.
- ²⁹ The distinction between partnership and consultants is related to the need to pay for the work done.
- ³⁰ Completed IEs are more likely to have been designed in the first part of the period of analysis. During interviews, many staff explained how they looked for support from external experts during the first operations that included an IE. Today they mainly use internal resources.
- ³¹ For example, the models of DIME and J-PAL show that strategic partnerships create the conditions to encourage creativity and specialization to develop new knowledge. DIME uses the network and reach of the World Bank to its advantage; it works with team leaders, country office representatives, governments, specialists in different sectors, and members of international specialized bodies to strengthen the IEs it produces and amplify their use and dissemination. J-PAL uses a carefully crafted academic network, embedded in the world’s best universities, to share knowledge and experience, define benchmarks, and keep the reputational quality of their work high.
- ³² These funds are charged to the borrower as part of loans. It is usually not possible to separate the resources allocated to monitoring from those allocated to IEs during project implementation.
- ³³ For example, the (now legacy) KCP products were to a large degree funded through TCs.
- ³⁴ All countries have at least one IE with collected data. Countries with more than one include Costa Rica and Trinidad & Tobago (2 each); Belize (4); Guatemala and Honduras (5 each); Nicaragua (6); Jamaica (7); El Salvador (8); Haiti and Paraguay (9 each); and Bolivia (11).
- ³⁵ Annex III provides more details on how these organizations do IE and on their costs.
- ³⁶ The loan amount does not include investment grants.
- ³⁷ This does not include reimbursable TCs or investment grants.
- ³⁸ This is in line with the strategy being set by IIC to guide their impact evaluations.
- ³⁹ Such a review process is the basis of the qualitative review of academic work in all academic journals. It is also the basis for selecting the research funded by all major public and private funding agencies throughout the world.
- ⁴⁰ CLEAR-LAC was selected through a competitive procurement process. A team of six researchers implemented the quality assessment instrument. Most of the participants were researchers from CIDE with economic and quantitative training and with previous experience in the implementation or review of IEs. The distribution of IEs for review among the participants was made considering the researcher’s sector expertise. A subset of the assessment questionnaires was peer-reviewed to promote the clarification of criteria to complete the instrument.

- ⁴¹ The framework used for this purpose is commonly used in the literature to identify basic characteristics of IE that provide compelling evidence to support what works and what does not in different fields; see Sherman et al. (1998), Halpern (2014), Gertler et al. (2016), and Karlan and Appel (2016).
- ⁴² This sample included all completed evaluations except for seven very poor-quality documents that OVE received, still in draft format, and one published in the *Quarterly Journal of Economics* that had already passed through a rigorous quality assessment.
- ⁴³ This is a stratified sample by department.
- ⁴⁴ A random stratified sample of 59 IE proposals was selected to do the quality assessment analysis. The analysis was also performed by CLEAR-LAC with a different version of the quality framework developed by OVE.
- ⁴⁵ The number of people reached by this training is likely to be lower, as the same person can participate in multiple activities.
- ⁴⁶ For details, see Annex VIII.
- ⁴⁷ A lexical analysis using the term “impact evaluation” (in both English and Spanish) was run through IADB documents—loan approval documents; and country strategies and technical notes—using MAXQDA software. This lexical analysis of IE citations in project documents is limited by the nature of its objective, since academic sources are known to be under-cited in political decisions, intervention designs, and project documents (Behrman, 2010).
- ⁴⁸ World Bank (2009).
- ⁴⁹ In 30% of the cases the Bank no longer had a relationship with anyone in the government; other factors were recent changes in the administration, loss of institutional memory about the evaluation, or the Bank specialist’s judgment that it was an inappropriate moment to contact the counterpart.
- ⁵⁰ In more than half of the cases the Bank seems to have put extra effort into training the executing agency.
- ⁵¹ With regard to the types of publications in which the IDB’s IEs are published, the dominant destination (36%) is IDB technical notes; sizable proportions are also published in peer-reviewed journals (13%), through research portals such as IDEAS or SSRN (13%), and on counterpart government websites (8%).
- ⁵² The collection was produced using Publish or Perish software produced by Anne-Wil Harzing of the University of Melbourne, which includes fields for author’s affiliation, year of publication, and key terms. Our search included any reference to the Bank (IDB, Inter-American Development Bank, or IADB), the term “impact evaluation,” and years of publication between 2006 and 2016. The software is based on Google Scholar and calculates the number of citations per article, as well as overall citation metrics for a search. It is like the traditionally used SCOPUS and Thompson ISI databases, but has a better coverage of non-English-language publications, and includes working papers and books in addition to academic articles from peer-reviewed journals. The same search procedure was run for the four official IDB languages.
- ⁵³ The citation analysis presented here was done in October 2016, and therefore analyzes data up to September 2016.
- ⁵⁴ This analysis was done using databases maintained by SCOPUS and Thompson-ISI. Similar analysis was done for the more recent years using Publish or Perish that serves as an interface to collect citation data from Microsoft Academic. The comparison between the IDB, World Bank, ADB and AfDB leads to the same conclusion.
- ⁵⁵ Annex VII provides details of the econometric method used and the complete analysis.
- ⁵⁶ This exercise is only indicative, providing a broad estimate of the savings generated by the average disbursements rates of IDB loans. These savings can vary significantly by sector, maturity, and country.
- ⁵⁷ The average amount of an investment loan is US\$83.9 million with IE and US\$92.7 million without IE. The current credit fee is 0.5%, and its monthly equivalent was applied to this example. When the amount used for the comparison is the same in both cases, the savings could go from 14% (US\$50 million) to 25% (US\$100 million).

- Banerjee, Abhijit, and Esther Duflo (2008). “The Experimental Approach to Development Economics.” Department of Economics and Abdul Latif Jameel Poverty Action Lab. Cambridge, MA: Massachusetts Institute of Technology.
- Bedoya, Guadalupe, Lucca Bittarello, Jonathan Davis, and Nikolas Mittag (2017). *Distributional Impact Analysis. Toolkit and Illustrations of Impacts Beyond the Average Treatment Effect*. Policy Research Working Paper 8139. Washington, DC: World Bank.
- Behrman, J. R. (2010). “The International Food Policy Research Institute (IFPRI) and the Mexican PROGRESA Anti-Poverty and Human Resource Investment Conditional Cash Program.” *World Development*, 38(10) :1473–1485.
- Blattman, Chris (2008). “Impact Evaluation 2.0.” Presentation to the Department for International Development, February 14, 2008. London: DFID.
- Cameron, D. B., A. Mishra, and A.N. Brown (2016). “The growth of impact evaluation for international development: How much have we learnt?” *Journal of Development Effectiveness* 8(1): 1-21.
- Center for Global Development (2006a). A Major Step Forward on Impact Evaluation. Available: <http://www.cgdev.org/content/article/detail/8207/>.
- _____ (2006b). “When Will We Ever Learn? Improving Lives Through Impact Evaluation.” Report of the Evaluation Gap Working Group. Washington, DC: Center for Global Development.
- Cunha, F., J. Heckman, and S. Schennach (2010). “Estimating the Technology of Cognitive and Non-Cognitive Skill Formation.” *Econometrica* 78(3): 883-931.
- Deaton, Angus (2009). Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development. NBER Working Paper No. 14690.
- Deaton, Angus (2010). “Instruments, Randomization, and Learning about Development.” *Journal of Economic Literature*, vol. 48 (2): 424-55.
- Dhaliwal, I., and C. Tulloch (2012). “From research to Policy: Using Evidence of Impact Evaluations to Inform Development Policy.” *Journal of Development Effectiveness* 4(4): 515-536.
- Duflo, E., and M. Kremer (2003). “Use of Randomization in the Evaluation of Development Effectiveness.” Paper prepared for the Conference on Evaluation and Development Effectiveness, 15–16 July, World Bank, Washington, DC.
- Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M. J. Vermeersch (2016). *Impact Evaluation in Practice*. Second edition. Washington, DC: Inter-American Development Bank and World Bank.

- Halpern, David (2014). *What Works? Evidence for Decision Makers*. London: Department for International Development.
- Imbens, Guido W., and Donald B. Rubin (2008). “Rubin Causal Model.” In *The New Palgrave Dictionary of Economics*, second edition, edited by Steven N. Durlauf and Lawrence E. Blume. New York: Palgrave.
- International Initiative for Impact Evaluation (2011). 3IE Impact Evaluation Practice: A Guide for Grantees. Available at www.3ieimpact.org/strategy/pdfs/3ie_impact_evaluation_practice.pdf.
- Karlan, Dean, and Jacob Appel (2011). *More than Good Intentions*. New York: Dutton Press.
- Karlan, Dean, and Jacob Appel (2016). *Failing in the Field: What We Can Learn When Field Research Goes Wrong*. Princeton: Princeton University Press.
- Legovini, Arianna, and Ferdinando_. Regalia (2001) “Targeted Human Development Programs: Investing in the Next Generation” Sustainable Development Department, Best Practices Series. Washington, DC: Inter-American Development Bank.
- Legovini, Arianna, Vincenzo Di Maro, and Caio Piza (2015). *Impact evaluation helps deliver development projects*. Policy Research Working Paper no. WPS 7157; Impact Evaluation series. Washington, DC: World Bank Group.
- Levine, Ruth (2010). Remarks before the U.S. Department of State Third Annual Conference on Program Evaluation (Washington, DC). June 9, 2010.
- Levy S. and E. Rodriguez (2004). Economic crisis, political transition, and poverty policy reform: Mexico’s PROGRESA-Oportunidades Program. Policy Dialogue Series. Washington, DC. IDB.
- Levy, Santiago, and Evelyne Rodríguez (2005). *Sin Herencia de Pobreza: El Programa Progres-Oportunidades de México*. Washington, DC: Inter-American Development Bank.
- Ludwig, Jens, Jeffrey R. Kling, and Sendhil Mullainathan (2011). “Mechanism Experiments and Policy Evaluations.” *Journal of Economic Perspectives* 25 (3): 17–38.
- OECD. (2006). Outline of Principles of Impact Evaluation. Available at <http://www.oecd.org/dac/evaluation/dcdndep/37671602.pdf>
- Office of Evaluation and Oversight (2002). “Analysis of Project Evaluability Year 2001” (RE-275). Washington, DC: Inter-American Development Bank.
- _____ (2014). “Evaluation of the Results of the Realignment” (RE-275). Washington, DC: Inter-American Development Bank.

- Pritchett, L. (2002). "It pays to be ignorant: A simple political economy of rigorous program evaluation." *The Journal of Policy Reform* 5(4): 251-269.
- Pritchett, Lant and Justin Sandefur (2013). "Context Matters for Size: Why External Validity Claims and Development Practice Don't Mix", Center for Global Development Working Paper 336.
- Ravallion, M. (2008). *Evaluation in the Practice of Development*. Policy Research Working Paper 4547, Development Research Group. Washington, DC: World Bank.
- Rothstein, A., J. Sutton, and M. Borenstein (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Chichester, England, and Hoboken, NJ: Wiley.
- Ruprah, I. J. (2008). "You Can Get It If You Really Want': Impact Evaluation Experience of the Office of Evaluation and Oversight of the Inter-American Development Bank." *IDB Bulletin* 39(1): pp. 23-35.
- Savedoff, W. D., R. Levine, and N. Birdsall (2006). *When Will We Ever Learn? Improving Lives through Impact Evaluation*. Washington, DC: Center for Global Development. Available: <http://www.cgdev.org/content/publications/detail/7973>.
- Sherman, Lawrence W.; Denise C. Gottfredson, Doris L. MacKenzie, John Eck, Peter Reuter, and Shawn D. Bushway (1998). *Preventing Crime: What Works, What Doesn't, What's Promising*. Washington, DC: National Institute of Justice.
- Sutcliffe, Sophie and J. Court (2005): *Evidence-Based Policymaking: What is it? How does it work? What relevance for developing countries?*. ODI Working Paper
- The Network of Networks on Impact Evaluation. (2009) *Impact Evaluations and Development*. Nonie Guidance on Impact Evaluation. Independent Evaluation Group. World Bank, Washington, DC.
- Vivalt (2016). *How Much Can We Generalize from Impact Evaluations?* Unpublished paper
- Weyrauch, V., and G. Díaz Langou (2013). "Sound expectations: From impact evaluations to policy change." *Journal of Development Effectiveness* 5(3): 269-304.
- World Bank (2009). "Making Smart Policy: Using Impact Evaluation for Policy Making—Case Studies on Evaluations that Influenced Policy." *Doing Impact Evaluation* #14. Washington, DC: World Bank.