Impact Evaluation and the Project Cycle

Thematic Group on Poverty Impact Analysis, Monitoring and Evaluation

Impact Evaluation and the Project Cycle

May 2006

Acknowledgement

Introduction

The goal of an impact evaluation is to attribute impacts to a project and to that project alone. To do this, a comparison group is needed to measure what would have happened to the project beneficiaries had the project not taken place. The process of identifying this group, collecting the needed data, and conducting the relevant analysis requires a lot of careful, thorough thought and planning.

A good impact evaluation provides the basis for sound policy making. It helps us to understand whether or not the project has had an impact, how large the impact is, and who has benefited (or not). In addition to providing hard evidence which can be used to weigh and justify policy priorities, impact evaluation can also be used as a managing-by-results tool. As the evaluation progresses side by side with the project, the evaluation can be used to test features of the project to modify design and improve effectiveness over time. Among other things, impact evaluations can help policy makers examine the effect of a pilot, compare different delivery modes, and examine the impact of the project for different populations. Overall, impact evaluation allows us to learn which projects work in which contexts, and use these lessons to inform the next generation of policies not only in the country of concern, but also across countries. Finally, the exercise of carrying out an impact evaluation helps to build and sustain national capacities for evidence-based policy making.

There is no standard approach to conducting an impact evaluation. Each evaluation has to be tailored to the specific project, country and institutional context and the actors involved. This said, there are some general questions and actions that bear on the success of any impact evaluation. ***The purpose of this document is to help task team leaders (TTLs) know what to do at different stages in the project cycle to secure a successful evaluation.***

Before turning to specific recommendations for the integration of evaluation into the project cycle, it is worth considering some overarching themes that should be addressed when planning and implementing an impact evaluation.

Integrating the impact evaluation with the project

A good impact evaluation is not a free standing exercise. Rather, it should be closely linked to the project. This is a two sided process. On one side, the lead evaluator (LE) should become intimately familiar with the project, the country and institutional context, the design options that are being considered, and the details of the roll-out and execution. On the other side, the TTL and client need to buy into the logic of the evaluation to understand what project design and implementation elements are critical for carrying out an evaluation that will contribute to improving the success of the project. Ideally, these two sides should come together as the TTL, lead evaluator and counterparts work through the design choices of the project to identify which of these choices need to be tested.

Relevance

For an evaluation to be relevant, it must be designed to respond to the policy questions that are of importance to the clients. Clarifying early what it is that the client wants to learn and designing the evaluation to that end will go some way to ensure that the recommendations of the evaluation will feed into policy making. This requires structuring the evaluation in such a way as to be able to answer the questions at hand and answer them at the time when feedback can be incorporated. For Bank projects, recommendations might be timed with project mid-term review and closing. For governments, they might be timed with budget discussions, PRSP preparations, or PRSP progress reports depending on context and needs. This leads to the next principle.

Government Ownership

Government ownership of the project is central to success. Government involvement is critical for a) identifying the relevant policy questions, b) ensuring the internal integrity of the evaluation -- helping to make sure that the evaluation keeps up with the reality of implementation, and c) to make sure results are incorporated in future policy choices and used to inform the allocation of resources across programs/sectors. In order to ensure this involvement, TTL and LE will work with the government on evaluation options, methods and execution from the start. This will be an opportunity to further develop domestic capacity for evidence-based policymaking.

Flexibility and adaptability

The evaluation must be tailored to the specific project. It should adapt to context and, whenever possible, capitalize upon planned changes in the surrounding environment. This requires cooperation between the TTL, lead evaluator, and the government counterparts.

To secure flexibility, the evaluation must be planned as early as possible (which is earlier than you think). While there may be a significant amount of uncertainty regarding the project activities at the early stages, by understanding the potentially large set of design options, the evaluation team will be in a better position to identify evaluation options.

Flexibility also requires close attention to the political environment. The evaluation team needs to identify and communicate with relevant constituencies who may support or oppose the evaluation and keep an eye on how these constituencies may shift during the life of the evaluation.

The tension of timing

There is a potential for tension between the objectives of the evaluation and the timeframe of the project. While some of the impacts should be visible within a short time frame, others may only be observed after project completion. Building an adequate results framework that recognizes the issue of timing will help design a realistic

evaluation. Selection of outcomes and indicators that reflect the timing of the project and the evaluation will be necessary. For longer term outcomes, a financial strategy that allows for evaluation beyond project completion will be necessary.

Guide to this document

The purpose of this document is to provide guidance for Bank staff on the process of designing and executing impact evaluations. This document is not a technical guide on impact evaluation methods[1], but rather presents practical considerations for the management of an impact evaluation.

This document is structured around three periods in the life of the project. The first period covers identification through the writing of the PCN. The second period spans preparation through appraisal. The third period is from appraisal through the completion of the impact evaluation (which may be later then the end of the project).

Each period of the project cycle is divided into six sections to help identify different types of activity, namely:

- Evaluation activities –the core issues surrounding the evaluation design and implementation
- Building and maintaining constituencies –the dialogue with the government and Bank management
- Project considerations –project logistical considerations that relate to the evaluation
- Data –issues related to building the data needed for the evaluation
- Financial resources – considerations and no easy answers for securing funding
- Support – resources available to TTL to assist in the evaluation process

Tasks are divided among the members of an impact evaluation team (IET). This team will be comprised of:

- The TTL who is responsible for project design on behalf of the Bank (TTL);
- A lead evaluator (LE) who is responsible for coordinating the design and implementation of the evaluation and analysis of the results. LE is often shorthand for a team of evaluators and data collectors; and
- The government counterparts (GC).

---

[1] For more on methods, see IE website, which lists a number of useful papers:
www.worldbank.org/impactevaluation

I. Identification to the PCN

A. Evaluation Activities

Question the reasons for doing an evaluation of the project (TTL). The TTL should question whether the evaluation will be useful, how to justify the expenditures, whether sufficient evidence already exists to gauge the effectiveness of this type of project, and whether there is reason to believe that the evaluation can improve the delivery of this or future projects in this country and sector.

Identify the policy questions that the evaluation might address (TTL/GC/LE). At this point the specifics of the project might not be defined. However, it is crucial to start identifying the policy questions that this work will try to address and thus useful at this stage to start thinking about for which components of the project impact evaluation is relevant and feasible. Indeed, it is quite possible that the impact evaluation will not be able to examine all interventions under a given project but instead focus on a set of components of interest. The identification of policy questions and which components to evaluate is an iterative process that will go hand-in-hand with project design.

Build the evaluation team (TTL). This with require identifying an international and/or local team of evaluators. Special attention should be placed in the selection of the lead evaluator who will ensure quality of the final product. The key skill a lead evaluator needs to have is familiarity with, and ability to apply, a range of impact evaluation methodologies. This person may come from inside or outside the Bank, depending on the skills needed and availability. The lead evaluator will need to supervise a team comprised of data collectors and (possibly) other evaluators. The TTL would then organize a discussion between government counterparts, project team and researchers to kick off the dialogue.

Think about how the team will work together (TTL). This will include making sure that the LE participates in exploratory and follow-up discussions with the project team and the government counterpart; that the government assigns one person who will liaise with the rest of the team and facilitate activities on the ground; and that the likely data collection agency is involved to ensure sharing of available data sources and that data collection activities are made consistent with national plans. This is also a good point to start thinking about building capacity for impact evaluation in the government and local academia and making a plan for it.

Write a paragraph for inclusion in the PCN which identifies the motivation for the impact evaluation (TTL).

B. Building and maintaining constituents

Identify constituencies and help build broad-based support for the impact evaluation. Identify a likely champion from the government side (TTL/GC). (See Appendix 1 for some suggestions on strategies to build government support)

Inform management (TTL). Make sure country and sector directors are informed and involved in the early stages of the evaluation. This will help when it comes time to identify financial resources, dialogue with government, and ensure sustained support as the project continues.

C. Project considerations

Draw link between the impact evaluation and the CAS (TTL). As the Bank has moved to results-based CASs, thought should be given to the issue of how the results of this evaluation will provide evidence on existing CAS results and help inform CAS priorities in the future.

Understand how the evaluation can build on the existing project and sector knowledge base and feed into future projects in this sector and country (TTL/LE). This requires a view as to what is known to work and where there are knowledge gaps. In addition to a literature review, it is helpful to organize a meeting of practitioners around these questions.

D. Data

Explore what data exists and might be relevant for use in the evaluation (LE). This requires discussions with the agencies of the national statistical system to identify existing data sources and future data collection plans. It also may be useful to look at what data academics and NGOS are collecting.

Start identifying additional data collection needs (LE). In the end, data and financial resources are inter-related but this is a good time to start thinking about possible funding/data scenarios (LE/TTL).

E. Financial resources

Identify likely sources of funding (TTL/LE/GC). Start thinking about how the evaluation will be funded and what the size of the total resources available is likely to be. The scope of the evaluation will depend in part by the available budget envelope. Conversely, an evaluation of significant scope might attract additional funding. Source might be internal to the Bank and external such as government and donors. PHRD grants

might be a good source for preparatory work including support for the LE to develop the analytical framework and survey instruments; and collection of baseline data.

Define under what contractual arrangements to hire the LE (TTL/GC). Keep in mind that, under Bank rules, if the LE is contracted through Bank budget funds, he/she will be ineligible to compete for government contracts for implementation of the evaluation. This means that the LE might be hired either by either the Bank or the Government for the length of the assignment (design and implementation).

F. Support

Identify the contact person for impact evaluation in your unit/network (TTL). This person may be able to refer you to consultants, financial resources and/or provide specific guidance for the evaluation.

Identify consultants (TTL). The Impact Evaluation Thematic Group (TG) has a roster of evaluation consultants [http://ieexperts.worldbank.org]. This searchable roster has a list of internationally recognized evaluation consultants. It includes their CV as well as the contact information.

Understand the methods and learn about existing evaluations in your project area and country. For information on impact evaluation methodology see the TG website [www.worldbank.org/impactevaluation]. For more information about prior evaluations, the TG maintains a searchable (by sector and country) database of completed impact evaluations at [http://www1.worldbank.org/prem/poverty/ie/evaluationdb.htm] which includes a summary of the results as well as methods.

Ask for project-specific support. For support on the design of an impact evaluation, on-demand impact evaluation clinics are offered by PRMPR and HDNVP. These clinics bring together sector and impact evaluation experts to work with the project team. Although you may not be ready for a clinic at this stage, start thinking about this possibility now – these clinics are best held as soon as a rough notion of the project activities and intended beneficiaries has been developed. For more information see the registration website: http://www-wbweb.worldbank.org/prem/prmpo/ie/request_ieclinic.cfm.

II. Preparation through appraisal

A. Evaluation Activities

*Identify design features of the project that will affect evaluation design (TTL/GC).*These include:

- Project target population and rules of selection. For the purpose of the evaluation, it is important for the project to spell out the target population for the intervention and how the ultimate beneficiaries will be selected. This will provide the evaluator with the tools for developing the initial strategy to identify project impacts.

- Roll out plan. It will also be important to start thinking about how the implementation will be rolled out for the project and its components. This will not only provide the evaluation with a workable framework for timing data collection activities, but may also help in the development of a strategy to identify project impacts.

Start to develop the identification strategy (i.e. how to identify the impact of the project separately from changes due to other causes) (LE). The LE will select the sample of units (households, firms, communities, depending on the project) that will best serve as a comparison against which to measure results in the sample selected into the project. The comparison group will be, at least, observationally equivalent to the group selected into the project, and, at best, equivalent on both observed and unobserved characteristics. The rigor with which the comparison group is selected at the beginning of the evaluation will determine much of the quality and reliability of the impact estimates obtained during analysis.

The way in which the project is implemented will affect the ability of the LE to develop a valid identification strategy. At this stage it is worth examining the possibility of project refinements that will enhance the evaluation – for example the way in which the project is phased in, clarification of eligibility criteria, and procedure for selection of pilot areas. The point here is not to change the development objectives of the project or to change the intended beneficiaries but to explore marginal changes to project design that will improve the quality of the evaluation. Optimally, identification strategy and project selection are developed in a way as to meet both project and evaluation objectives.

In some cases, the identification strategy might be able to capitalize on the phase in of the project and identify the comparison group as the group that will benefit from the project in phase II or III. The evaluation will then be done within the time framework allowed by the last phase of the project and will not be designed to look at long term effects. This strategy is particularly attractive when the project plans national coverage by the end of the project period. The objective of the evaluation will then be to confirm the effectiveness of the intervention before full scale up. As indicated above, the TTL and

GC might need to work with the LE to adjust the selection of units for phase I and phase II roll out to support the evaluation objectives.

Develop the results framework for the project (TTL, LE). This activity will clarify the results chain for the project, identify the outcomes of interest and the indicators best suited to measure changes in those outcomes, and the expected time horizon for changes in those outcomes. This will provide the LE with some of the information necessary to develop the survey questionnaire and schedule data collection.

Develop a view regarding which features of the project should be tested early on (TTL, GC, LE). Early testing of project features (say 6 months to 1 year) can provide the team with the information needed to adjust the project early on in the direction most likely to deliver success. Features might include alternative modes of delivery (e.g. school-based vs. vertical delivery), alternative packages of outputs (e.g. awareness campaigns vs. legal services), or different pricing schemes (e.g. alternative subsidy levels). This approach will require a different timing for the data collection and a better understanding of short-term impacts or proxies, than an evaluation solely designed to measure project impacts at closing.

Identify other interventions in the project area (TTL, GC). It is important to take stock of **all** existing and planned programs in the areas that will receive the project and in those that you plan to use for comparison. These include not only Bank projects, but government, NGO and other donor activities. If they are not taken into account, they can pose a threat to the validity of some of the evaluation. When taken into account, they may open the door for testing synergy of different interventions.

Think about risks to the impact evaluation. These include threats to validity, such as existing projects in the program area, sample attrition, threats stemming from problems or unexpected changes in project implementation (e.g. delays in disbursement, procurement problems, lack of adherence of eligibility criteria), and threats to the execution of the evaluation (e.g. potential political/administration changes, funding uncertainty).

B. Building and maintaining constituents

Maintain dialogue with government, build a broad base of support, and include stakeholders (TTL/LE). This will be critical in sustaining government support for carrying out the evaluation. Issues for discussion include:

- Tradeoffs between rigor and feasibility (see discussion above on identification strategy), i.e., discuss the most rigorous (in terms of attribution of impacts) design that is feasible – not only administratively but also politically. This is a critical step for reaching agreement with government on the design of the evaluation and important in order to ensure government support for the integrity of the evaluation design (IET).

- Dissemination plan. The plan should be agreed with government counterparts to maximize the likelihood that evaluation results will be applied to policy development. It will include an agreement on how and when interim and final results will be disseminated and fed back into the policy cycle. Particular attention should be placed on project and policy cycles to identify entry points for use of evaluation results (e.g. midpoint and closing, for project; sector reporting, CGs, MTEF, budget, for government). It should include the explicit identification of the mechanisms and processes for feeding the results back into policy and program design. Dissemination of results to the treatment and comparison groups might be considered (IET).

Build capacity (TTL/LE). Use collaboration with government officials and partnerships with local academics to build local capacity for impact evaluation. More activities can be implemented with the help of your regional/sectoral IE contact or the IE TG (training, workshops, clinics). The objective is to build institutional buy-in for the culture of evaluation and managing-by-results.

C. Project considerations

Integrate the impact evaluation into the project as a mechanism to increase the effectiveness of project implementation. For example, different stages of data collection will provide inputs to the project. Baseline data may be used to improve targeting efficiency. Early evaluation results can be used to motivate a change in project design at the time of mid-term review. (IET)

Formally initiate an IE activity (TTL). The IE code in SAP is a way of formalizing evaluation activities. The IE code recognized the evaluation as a separate AAA product. See Appendix 2 for further details about the code and how to use it. (TTL)

Write up a concept note for the impact evaluation (LE).

Identify peer reviewers (TTL/CE). Ideally these should include someone well versed in impact evaluation methods as well as someone familiar with the sector (TTL/LE).

Have a meeting to review the concept note for the impact evaluation, if applicable (IET).

Work the impact evaluation into Annex 3 of the project document (TTL/LE).

Include the impact evaluation in the Quality Enhancement Review (TTL).

D. Data

Explore existing sources of data and planned data collection activities to identify data that might address impact evaluation needs (LE).

- Include:
  - o Censuses
  - o Surveys (household, firms, facility, etc)
  - o Administrative data
  - o Data from the project monitoring system

- Record data periodicity (including future plans), quality, variables covered and sampling frame and sample size.

Investigate synergies with other projects to combine evaluations or at least evaluation data collection efforts (LE, TTL). It may be possible to combine the data collection efforts for multiple projects, e.g. using the same survey instruments and data collection infrastructure. Collaboration can improve efficiency in the use of local resources (e.g. statistical agencies), secure savings on survey costs, and ensure that other activities are properly accounted for in the evaluation; but collaboration can be costly in terms of coordination between projects.

Develop a data strategy for the impact evaluation (TTL/LE). In most cases, evaluation will require data collection. The plan will include:

- The timing of the data relative to the implementation of the project (e.g. when was the baseline collected/when should it be collected?)

- The variables needed. These need to cover not only the impacts of interest but potential controls (e.g. demographics, consumption) and (if applicable) instrumental variables. These will provide the basis for the questionnaire design if new data is to be collected.

- The size of the (feasible and sufficient) sample and the sample frame

- Integration of data from other sources

- Existing data collection efforts on which the new data collection could piggy back (if relevant). This could include adding a module or more observations to an existing planned survey.

E. Financial resources

Develop and finalize a budget and plans for the financing of the impact evaluation (TTL/IET). The costs should cover LE costs (from design to analysis),

supervision costs and data collection costs. Financing may include BB, trust fund, donor funds, research funds and project funds. Bank implemented impact evaluations will use the IE code in SAP to identify the product. Government implemented IE will have costs rolled into the project budget. Hybrid solutions may have the Bank finance the analysis and the government finance the data collection. If the evaluation is to be included in the project budget, you need to make sure to identify a way to pay for any data collection that falls outside of the project effective dates. Appendix 3 provides a list of the activities with some notional costs and some options for financing.

F. Support for the evaluation

For more information on surveys, data sources, and overall statistical capacity for a given country see the DECDG website (type data in an intranet browser). Useful pages include the country statistical information pages and the Development Data Platform, among others.

Schedule an impact evaluation clinic. Clinics bring together a group of experts to provide on-demand and project-specific support for the design of an impact evaluation. For more information see the registration page:
http://www-wbweb.worldbank.org/prem/prmpo/ie/request_ieclinic.cfm.

III. Negotiations to completion…and beyond

A. Evaluation Activities

While the evaluation has been designed by now (hopefully), it is critical that the TTL and LE monitor project <u>implementation</u> with an eye towards threats to the validity of the evaluation design. These may include delays to implementation that render the baseline obsolete, changes in program targeting, and/or major shifts in the project activities. This should also extend to paying attention to monitoring project <u>effects</u> (through the monitoring system and observation) for unintended consequences such as spillover effects on the comparison group. Not all these are threats to the evaluation. Some may present an opportunity to redesign the evaluation, increase its scope and effectiveness, and/or reduce data costs (TTL/LE).

Monitor the evaluation implementation (LE with input from the TTL). Make sure that tasks are completed on time and that analysis starts on schedule.

Start analysis as data becomes available and feed results into (LE). Use data and results of the analysis to inform changes in the project (mid-term) and project assessment at completion (ICR). Feed project changes back into the evaluation design as appropriate.

Disseminate results to inform the next generation of projects as well as government policy development (TTL).

Continue collaboration with government and local researchers to build capacity for impact evaluation (TTL/LE).

B. Building and maintaining constituents

Present evaluation results to government first before releasing them to a wider audience (TTL/LE). This will make sure that sensitive issues will be presented to the wider audience in a politically sensible fashion. Governments should not be threatened by evaluation results but given a chance to use them to improve government effectiveness. The approach will help build support for policy follow up on the evaluation results as well as avoiding political problems and backlash.

Preview results with Bank management, as well (TTL/LE).

Work with government closely on the dissemination of the results (IET). Make sure message is not distorted. Make sure caveats are clear and often repeated.

C. Project considerations

Involve and inform the local project implementation unit and PIU person responsible for monitoring and evaluation (TTL/LE).

Put in place arrangements to procure the impact evaluation work and fund it on time (TTL).

Use early/mid-term evaluation results for redesign of the later phases of the project (TTL).

Include the evaluation results in the ICR (TTL).

Feed evaluation results into the CAS (TTL).

D. Data

Implement data collection (IET):

- If this involves collecting new data: Carry out the baseline before implementation (note that this may or may not be before project effectiveness). Take steps to ensure data quality, including an explicit discussion prior to the baseline on mechanisms to ensure data quality (IET). The TTL and LE should closely monitor data quality throughout the collection of the baseline, data entry and any subsequent data collection.

- If this involves using existing data sources: the TTL and/or the LE needs to liaise with the agencies collecting the data to make sure that data gathering plans are proceeding on or close to schedule.

- If this involves collaborative/piggybacking data gathering (this potentially applies to both cases above): the TTL needs to maintain contact with the other organizations involved to avoid unexpected problems. In cases where this coordinated effort involves a significant number of actors, it may be useful to establish some sort of forum for this to take place.

E. Financial resources

Make sure evaluation funding is sufficient, raise additional funds as needed (TTL). The TTL should keep in touch with the LE during analysis to make sure that there are sufficient resources for the analysis.

F. Support for the evaluation

Ask for support from the IE contact person in your region/sector as needed.

Use web resources and IE Thematic Group resources as needed.
http://www1.worldbank.org/prem/poverty/ie/evaluationdb.htm provides examples of similar work in the country/sector if it exists.

Appendix 1
Some suggested ideas for increasing government buy in to an impact evaluation

In this appendix we provide some ideas and suggestions on strengthening the constituency for an impact evaluation. For a more detailed discussion of one team's experience with this, see the forthcoming note in this series which discusses experiences in Argentina.

- Work with the government counterpart from the beginning. Spend a significant amount of time explaining the motivation/usefulness and methods.

- Understand who the client(s) are and their needs. Develop an impact evaluation which answers the clients' priority questions.

- Understand the other interested parties – their position on the issues around the evaluation, how they might affect the evaluation, and how they may react to the results. Spend time with these groups explaining the motivation/usefulness and methods.

- Identify a champion within government. This advocate can help with arguing for the evaluation initially, making sure it is well executed, and helping to ensure that the results get used.

- Get local researchers involved. A good place to start are those involved in the PRSP, they are likely to have government access and understand the issues and context

- Building impact evaluation capacity (especially in terms of analytical work) within government will not only provide benefits in terms of evaluation quality (e.g. a better understanding of the institutional context and implementation) but will also help build a constituency for future evaluations.

- If possible, get high profile Bank involvement to support this in dialogue with the government.

- Facilitate access to high quality human resources – make it as easy as possible for the counterparts to deliver a quality impact evaluation. If you cannot identify the resources on your own, refer to some of the sources in the support sections above.

- Do some homework on costs and funding options before talking with government about the evaluation. Think about possible positive spillovers that the evaluation will generate (e.g. data on certain populations of interest).

- If applicable, provide interim result to the client(s). In addition to keeping the client up to date on the direction of the evaluation, these can be used for mid-term policy corrections if they are warranted.

  When the results are ready…

- Make caveats clear up-front and repeat – this will be important in ensuring that results are interpreted correctly.

- Timing of release is important. Understand the domestic political/decision-making cycle and time the release of your results in order to maximize the impact on policy.

- Don't surprise the key players. It's advisable to give them a preview of the results before broad dissemination.

- Provide constructive policy guidance. Identifying the magnitude of the impact and the effects on different groups is only the first step. Discussing the implications for policies that stem directly from the evaluation results will assist government in identifying the best ways to put the evaluation results into action.

- Communicate in different forms (e.g. technical, less technical) to different audiences, as much as possible both inside and outside government.

Appendix 2
Making an impact evaluation a Bank product

Note that this is a reprint of the DEC-OPCS "Implementing Impact Evaluations at the World Bank: Guidance Note"

As part of the results agenda, the Bank has made important efforts to expand and deepen its impact evaluation effort. Through the Development Impact Evaluation (DIME) initiative it has provided coordination and facilitation to units and staff throughout the Bank engaged in impact evaluation activities. Recognizing this new trend, OPCS established IE as a new product line. This note provides guidance to staff on the implementation of this new product line.

Background

Enhanced emphasis on monitoring and evaluation of programs constitutes a key aspect of results-based management approaches to development assistance. This is a multi-faceted challenge involving a range of activities: from monitoring and assessing progress in the implementation of programs, to measuring changes in outcomes and evaluating the impact of specific interventions on those outcomes. The World Bank is taking steps to expand and improve its own focus on these monitoring and evaluation activities.

Within the toolkit of evaluation approaches, impact evaluations play an important role. Their goal is to assess the specific outcomes attributable to a particular intervention (e.g. the increase in student learning resulting from a change in teacher hiring arrangements or the higher incomes among micro-entrepreneurs resulting from improved access to credit). They do so by using a counterfactual that represents the hypothetical state the beneficiaries would have experienced without the intervention. From that point of view, impact evaluations are an essential instrument to test the validity of specific approaches to addressing development challenges (e.g. reducing infant mortality or increasing productivity of poor farmers). They provide a powerful instrument to determine 'what works and what does not work' and, thus, constitute a fundamental means to learn about effective development interventions. At the same time, particularly when conducted in comparable and consistent ways across countries, impact evaluations can provide the necessary benchmarks for program design and monitoring.

In the past, such evaluations were constrained by the lack of data and the technical challenges of developing a counterfactual. Over the past few years though, some significant improvements in both these areas have made impact evaluations easier to implement on a systematic basis: micro data gathered through household surveys or demographic and health surveys are more available and a range of evaluation techniques have been developed to construct the counterfactual – from randomized experiments to quasi experimental techniques.

Box 1: Evaluation Methods

There are two basic approaches to constructing the counterfactual. Experimental designs (also known as randomized control designs) construct the counterfactual through the random selection of treatment and control groups. Given appropriate sample sizes, the process of random selection ensures equivalence between treatment and control groups in both observable and unobservable characteristics. The other approach, quasi-experimental designs (also called non-experimental designs), relies on statistical models or design features (which sometimes generate 'natural', unplanned, experiments) to construct a counterfactual and include approaches such as regression discontinuity design, propensity score matching, and instrumental variables. In practice, quasi-experimental methods are much more common than randomized control designs.

Guidance on implementing IE activities

Impact Evaluation (IE) is a new product line that has been established under the AAA umbrella. Other product lines in the AAA family are: Economic and Sector Work, Technical Assistance, Donor and Aid Coordination, Research Services, and the World Development Report.

To qualify as an IE, an activity must meet all of the following criteria:

(i) Involve empirical work to measure the effects on a set of key outcomes of a development intervention (which can be associated for example with a government or NGO run program/project or one of its components) relative to a well-specified counterfactual (i.e. what would have been the evolution of the outcomes without the intervention). The intervention to be evaluated does not need to be receiving direct World Bank financial support to qualify for a Bank supported impact evaluation.

(ii) Be aimed at informing an external client. External clients are typically the Bank's borrowers, but they also include other member countries, donors, nongovernmental organizations, private sector organizations in developing countries, or other members of the development community. External clients are not necessarily based in the country where the project is taking place.

(iii) Lead to a free-standing report even when it is linked to the preparation and/or supervision of a Bank-financed project/program. The report must be initiated by a specific Bank unit.

There are important differences between IE and other existing Bank products:

- Difference between ESW and IE: ESW is an analysis undertaken specifically with the intent of influencing an external client's policies and programs. On the other hand, an IE has the more limited objective of establish, through a rigorous analysis, the changes in outcomes that can be attributed to a specific intervention (project, program, or policy). These results could eventually lead to policy recommendations (through another Bank instrument) for the country or for other external clients, but is not the objective of the IE.

- Difference between IE and TA: In a non-lending TA, the Bank can support an impact evaluation conducted by a government, providing only technical advice on how to conduct certain components of an impact evaluation or building country capacity to conduct impact evaluations. Under an IE-coded activity, the Bank is actively engaged in all aspects of the IE, and delivers a final output for which it takes responsibility even when it is jointly done with other agencies (including Governments).

- Difference between IE and IEG's Evaluations: IEG's primary responsibility is to evaluate Bank's operations ex post along many dimensions – e.g. the relevance of each operation and its alignment within a country's development strategy, the extent to which the objectives of the operations have been achieved and are sustainable. An IE seeks to evaluate the impact of an intervention on the basis of a sound counterfactual, no matter if the intervention is supported financially by a Bank or not.

The processing of IE is expected to include the following steps:

(i) Creating the Activity Initiation Summary (AIS) in SAP. The designated managers (typically the Country Director and Sector Manager) approve the activity and the VPU Release Authority releases the funds. At this time, the actual AIS Sign-off date is automatically populated in SAP. In the Activity Initiation Summary (AIS), TTLs must select at least one Development objective (DO) which represents the goal in undertaking the particular IE and the related result indicator(s), which specify the interim outcomes associated with the development objective. In general, each impact evaluation needs a separate AIS. However, where clustering of several evaluations is justified (e.g. evaluation of different components of one program, or evaluation of impact on various outcomes), they can be grouped under single AIS. An IE is considered a joint product if (a) its preparation involves substantial financial or in-kind contribution, covering 10 percent or more of the cost of the output, from at least one multilateral or bilateral donor,[2] and (b) that agency accepts the product as a joint output with the Bank, as signaled clearly in the concept paper and the eventual report. The team leader should record the nature of the product by filling in the "joint AAA" flag in the AIS/AUS. OPCRX will regularly review all IE products at AIS/AUS stage and check that the activities are properly coded. The OPCRX team might contact the task team leader to request further information on the IE when necessary.

(ii) Prepare Concept Note: In general, a Concept Note covers: (a) the rationale; (b) objectives and methodology; (c) the expected impact of the work; (d) team composition and resources (Bank, partners, client, including Trust Funds); (e) the dissemination/follow-up plan, and (f) the timetable.

---

[2] Substantial support from a trust fund does not automatically qualify the product as a joint IE, unless the trust fund donor also accepts the output as a joint product. This support is signaled on the cover page or elsewhere in the final report.

(iii) Peer Review/Quality Enhancement. IE activity requires a peer review/quality enhancement. A list of recommended peer reviewers will be available to staff. VPUs are expected to define the specifics of the quality enhancement mechanisms (i.e. nature of the review process for concept papers and final evaluation reports and specific responsibilities) as for any AAA product.

(iv) Record Key Milestones. TTLs must enter the actual milestone *Activity Implementation Start/Concept Note* (i.e., the date when the team initiates the actual development of the product or the date when the concept note is reviewed or approved) in the Activity Update Summary (AUS) in SAP. They must also update the actual milestone "Delivery to Client" and "Final Delivery" (i.e., the date the final evaluation report is approved by the designated manager and submitted for publication).

(v) Publication and Dissemination. Upon final delivery, the evaluation report is submitted for publication in the World Bank's Development Impact Evaluation Working Paper Series with full recognition to the originating unit and team. This new series will be coordinated by the office of the Bank's Chief Economist (DECVP) and will have its own advisory committee that will be responsible for reviewing the methodological soundness of the IE before authorizing its publication. Dissemination of the results of the evaluation report remains the responsibility of TTLs who is required to file in IRIS all the dissemination material (e.g. list of participants, presentations, agenda, talking points, minutes of the meeting, etc).

(vi) Create the Activity Completions Summary (ACS) in SAP. ACS procedures require the task team leader to submit an ACS in SAP for the IE. The ACS should be completed within six months of delivery to client as long as all tasks associated with the activity are complete. The designated managers approve the ACS and the actual ACS date is automatically populated in SAP. In the Activity Completion Summary (ACS), TTLs must rate the result indicator(s) for each selected DO, reflecting the extent to which it has been achieved (fully, largely, partially, no).

(vii) Resources available to staff involved in IE activities. Staff involved in IE activities have access to a series of support initiatives: (i) a roster of recommended peer reviewers; (ii) a roster of external consultants with expertise in impact evaluation; (iii) an interactive database of impact evaluations of Bank-supported projects, covering methods and results, (iv) learning resources (including technical guidance materials, sector specific evaluation methods notes, training material, and other relevant resources) and (v) evaluation clinics. All these are accessible through a website (www.worldbank.org/impactevaluation).

Appendix 3
Illustrative costs of an impact evaluation

This annex provides some notional figures for the various activities that an impact evaluation may involve. Ultimately, all costs and especially those for data collection are going to depend on the impact evaluation method used. In turn, it is also likely that the available funds will affect the choice of method. Furthermore, data collection and consultant costs can vary significantly across countries. These figures are meant to be purely illustrative and individual experiences will vary.

Some basic components will be:

1. Lead Evaluator (needs to be the same person from pre-preparation to analysis after completion). This individual will be responsible for evaluation design, survey instrument design and data analysis. If the LE is external to the bank: For a senior researcher around $50-100,000 for professional fees plus travel (travel estimates of at least 3 trips, one of which will be during preparation). If the researcher is internal to the Bank: estimate about 5-15 staff weeks a year. *One issue to be considered is the procurement that regulates the hiring of the same consultants in preparation and during implementation. The lead evaluator **will** need to be involved in all stages of the project.*

2. Sample design expert, about $3000 professional fees plus travel (one trip). This cost will most likely occur during preparation.

3. Data collection: contract either the national statistical agency or a survey firm. The choice will depend on firm capacity and experience relative to the desired sample size. If the evaluation requires a separate survey, the cost will depend on the sample size, questionnaire length, and the geographical dispersion of the sample. Per interview costs can range from $50 to $150 per respondent per interview. For example, a sample size of 1000 households for two rounds of data collection will cost from $100,000 to $300,000. If it is possible to piggy-back on an existing survey, the cost will be lower than this.

4. Supervision through the length of the evaluation. This should be a locally based consultant who covers all aspects of the implementation of the evaluation. Costs here can range from $12,000 to $20,000 per year.

5. Supervision costs from Bank Budget funds will need to be budgeted for. This will likely include one trip per year, and may be combined with other supervision duties, depending on who in the Bank team is supervising the evaluation.

6. Dissemination of the interim and final results. This will entail two trips of either the counterparts to headquarters or of the evaluation team to the country.

Sources of financing:

Items 1 and 2 will require project preparation funds for all work that takes place before approval. Grant financing (or the use of the IE code within the BB) is necessary to fund the later stages of the lead evaluator's work (which are likely to extend beyond the life of the project).

Items 3 and 4 can be financed through project funds for activities within the project cycle. For data collection outside of the project cycle other government funds or grants should be used.

Item 6 will require grant financing or BB (possibly through the IE code).