USAID
FROM THE AMERICAN PEOPLE

ASSESSING THE EFFECTIVENESS OF ECONOMIC GROWTH PROGRAMS

PRIVATE SECTOR DEVELOPMENT IMPACT ASSESSMENT INITIATIVE

MAY 2010

This publication was produced for review by the United States Agency for International Development. It was prepared by DAI.

ASSESSING THE EFFECTIVENESS OF ECONOMIC GROWTH PROGRAMS

PRIVATE SECTOR DEVELOPMENT IMPACT ASSESSMENT INITIATIVE

Program Title:	**AMAP Business Development Services Knowledge and Practice Task Order**
Sponsoring USAID Office:	**USAID Office of Microenterprise Development**
Contract Number:	**GEG-I-800-02-00014-00, Task Order No. 01**
Contractor:	**DAI**
Date of Publication:	**May 2010**
Author:	**Lucy Creevey, Jeanne Downing, Elizabeth Dunn, Zan Northrip, Don Snodgrass, and Amy Cogan Wares**

The authors' views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development or the United States Government.

CONTENTS

INTRODUCTION

The U.S. Agency for International Development (USAID) Private Sector Development Impact Assessment Initiative (PSD IAI)—created by USAID's Office of Microenterprise Development in 2003—was established to gauge the effectiveness of a "new generation" of economic growth (EG) programs that had recently emerged with funding from USAID and other donors. These new-generation programs promised to achieve economic growth and poverty reduction through a more complex and holistic set of interventions than had been used in the past. Previously, specific types of EG interventions—such as business environment reform, trade facilitation, market access and linkages, rural and value chain finance, business service provision, and input market improvements—were usually addressed by distinct projects, each of which was presumed to provide the missing ingredient that could unlock growth. Today, project implementers look at private sector development more holistically. They acknowledge the existence of multiple, interrelated constraints and draw from an integrated set of intervention tools in an effort to foster broad-based growth and catalyze systemic economic change, usually within multiple value chains.[1]

Although this trend in EG programming has been documented, evidence on its effectiveness remains sparse. To some extent, this lack of evidence is attributable to the relatively recent emergence of the approach. But it is clear that the inherent nature of these programs creates significant challenges for impact assessment. New-generation EG projects combine multiple, diverse interventions and work through a broad range of market actors at all levels of the value chain, including agricultural input providers, farmers, managers of local product outlets, private sector trainers or extension agents, mid-level retail agents, credit providers, exporters, and trade groups. The projects are also concerned with the broader context of the value chain—the regional and global market in which the value chain operates and the government regulations that either constrain or facilitate its growth and development.

Why It's So Difficult

In a drug trial, evaluators contrast the outcomes of a treatment group that receives a specified amount of a drug with those of a control group that receives no drug at all. In a typical EG project, evaluators contrast the outcomes of a treatment group that participates in widely varying degrees of intensity in multiple types of interventions to those of a control group that is not specifically targeted but may also be exposed to the treatment through spillover effects or the similar efforts of other government or donor projects. Moreover, project implementers usually change their "treatments" as the project evolves, and administer certain economy-level treatments (such as legal reform) from which it is impossible to isolate the control group. The inherent nature of new-generation EG projects and the developing-country contexts in which they are implemented pose significant challenges to traditional impact assessment. The "labs" in which EG programming takes place are far from sterile and are often unstable as well.

Moreover, EG project implementers are encouraged to respond quickly to changing market conditions and to take advantage of new opportunities. Implementers seek out and collaborate with private sector partners identified as lead firms and catalysts for change, aiming to spread positive effects up and down the value chain. While implementation agility may contribute to the success of the intervention, it also creates difficult problems for impact assessment, which aims to isolate and establish the impact of a specific activity over time.

[1] http://apps.develebridge.net/amap/index.php/Chain_Analysis

Even in the face of evaluation challenges, donors and project implementers need useful, credible information on the effectiveness of their programs. This paper suggests a way forward. Based on lessons learned under the PSD IAI, we identify and describe specific features of EG programs that challenge established impact assessment methods. Next, we introduce the degrees of evidence framework, a set of principles that can inform the design of an EG effectiveness evaluation and guide the selection of evaluation methods. The degrees of evidence framework leads to the selection of a portfolio of evaluation methods, including established impact assessment that provides a strong case for attribution as well as complementary evaluation methods that create a preponderance of evidence about overall project effectiveness. Before concluding, we illustrate the use of the framework by applying it to the design of a typical EG project.

CHALLENGES TO ASSESSING IMPACTS OF EG PROJECTS

The need to assess the effectiveness of project interventions in improving value chain performance and, ultimately, the income and well-being of its actors has resulted in a serious effort to develop rigorous methods of assessing project impact. Researchers generally agree that rigorous impact assessment requires defining and presenting a plausible *counterfactual*—that is, an indication of what would have happened if the project had not taken place.[2] However, the nature of EG projects makes this standard difficult to attain, given that the effectiveness of the overall project is to be assessed. Based on recent experiences conducting impact assessments of EG projects in Kenya, Zambia, India, Brazil, and Azerbaijan, we have identified five major challenges inherent to evaluating EG projects. These challenges relate to 1) project complexity, 2) policy-level intervention, 3) project evolution, 4) spillover effects, and 5) participant identification. In addition, several other difficulties routinely affect survey-based field studies.

PROBLEMS INHERENT TO EG PROGRAMMING

Project complexity. Project complexity refers to the typically wide scope of EG projects. EG projects studied in Kenya and Brazil show how sheer complexity can undermine the possibility of comprehensive impact assessments based on valid counterfactuals. In Kenya, the impact assessment covered three tree fruits, parts of two USAID projects with different approaches and multiple interventions and five geographic areas. This made a complete evaluation of project activities directed at the three selected tree fruits extremely difficult. In Brazil, the project focused on three regional subsectors: beachwear in Salvador, Bahia; cashew nuts in Barreira, Ceará; and honey in Simplício Mendes, Piauí. Four intervention areas were implemented in each subsector: 1) brokering commercial linkages; 2) upgrading; 3) facilitating access to financial services; and 4) identifying barriers to sector competitiveness.[3] While it is theoretically possible to set up valid counterfactuals and thus demonstrate what would have happened if the project had not intervened, in practice it is difficult to conduct comprehensive impact evaluations of a project operating in multiple locations and multiple value chains where timing, conditions, and types of interventions are different.

Policy-level interventions. A second major evaluation problem for EG projects arises when projects implement interventions to generate changes in the business enabling environment, such as national,

---

[2] Don Snodgrass, "Assessing the Impact of Private Sector Development Projects," Impact Assessment Primer Series (IAPS, December 2006).

[3] *Ibid*., p. ix

sectoral, or regional policies and regulations, or bring about other types of change that affect the entire value chain, such as changes in inter-firm relationships and the value chain governance structure. It is impossible to prove attribution if what is being studied is project-recommended changes in government regulations and strategies. One can infer, if a new policy is adopted, that it may have led to whatever positive outcomes are observed, but research cannot prove this since everyone is subject to the new policy and no control group can be formed or observed. Thus, policy-level interventions pose especially difficult problems for assessment of project impact.

Project evolution. The continuous evolution of EG projects, the third major issue, is one of the most difficult for research to handle. Important strengths of EG projects are that they are frequently viewed as ongoing experiments and can adapt to changing circumstances as they develop. Adjustments are often made mid-course, based on the results of early intervention activities as monitored and observed in the field. The adaptations may involve discarding interventions that are not working and undertaking efforts to seize new opportunities that are identified in the course of program implementation.

Projects may also evolve to adjust to changes in relevant outside factors, such as market conditions, policies, and regulations. Intervention models, intervention partners, and even the basic intervention paradigm can all change over time.[4] In Zambia, the shifts in project focus and partners were extreme. That project supported interventions in smallholder cotton cultivation, smallholder cattle husbandry, and retail distribution of farm inputs to smallholders. The impact study was designed to assess impacts on the cotton and beef value chains, on-farm input distribution, and the welfare of the farmers who participated in the projects and their households. But, as stated in the summary final report,

> …Although a larger than needed sample size was taken at the baseline, unusually high attrition among program respondents due to geographic shifts in program activity affected the statistical significance of results and rendered some statistical methods less relevant. Of the three sectors, the shift in cotton sector activity was the most drastic, resulting in program suspension after the collapse of the international cotton market in 2006. Nevertheless, retail services also shifted to more dynamic areas, as did the beef program due to lack of participating veterinarians. The lack of baseline data for subsequent activity meant that it was not possible to measure household level impacts and poverty mitigation….[5]

What is particularly telling about this case is that the project initially had an explicit causal model that was developed in the early stages of the research, when the research team met in the field with project staff to work out exactly what interventions would be planned with what expected outputs and outcomes.[6] This causal model was carefully developed but needed modification as the project shifted.

Spillover effects. The fourth major problem is spillover, which is a built-in feature of EG projects. The goal of these projects is to spread change and innovation throughout the value chain as quickly as possible. But if this spread occurs as the project progresses, it means that the participant group and the control group may not be very different in business practices, changes in income, or other outcomes.

---

[4] For example, value chain projects have replaced business development services (BDS) projects in USAID programming.

[5] Gary Woller, Amy Wares, Marina Krivoshlikova, Claire Simon, and Jennefer Sebstad, "Assessing the Impact of USAID/Zambia's Production, Finance and Improved Technology (PROFIT) Project, Draft Summary Report, August 2009.

[6] The importance of the causal model for research is discussed in Jeanne Downing and Gary Woller, "Developing a Causal model for Private Sector Development Programs, Impact Assessment Primer Series #4, January 2007.

An example of spillover can be found in the Azerbaijan project, in which two principal components were training veterinarians and providing access to newer drugs. During project's second year, participating veterinarians were encouraged to share what had they learned in training with other veterinarians. At the same time, the new drugs were made available to all veterinarians, whether or not they were a part of the project. It became next to impossible to find a significant group of veterinarians who had not been exposed to the project intervention strategy for the follow up study. Thus, while the study results showed increases in the amounts and kinds of services offered to farmers raising animals and poultry, it was not possible to identify a non-participant group of veterinarians to use in measuring the project's impacts.[7]

Spillover also occurred in India, where the project intervened in the domestic fresh vegetable market to link small-scale vegetable farmers to a major supermarket buyer. This quickly drew the attention of other buyers, who entered the regions and established commercial relationships with farmers outside the project. While this successful demonstration effect was one of the project's objectives, it complicated the impact assessment at the farmer level.

Participant identification. The difficulty in identifying participants is closely related to the spillover problem. At issue here is distinguishing between the treatment and control groups for the purpose of demonstrating the counterfactual. To use the experimental approach, program participants—the farmers or artisans whom the program hopes to benefit—must be selected at random from a population of qualified potential beneficiaries before program implementation begins. In the quasi-experimental approach, researchers draw a random sample of farmers in an area where the project intends to work and regard members of this sample as the treatment group. Similarly, a random sample of farmers in an area where the project does not intend to work is drawn and regarded as the control group. In practice, however, members of the treatment group may participate in the project to varying degrees and in various forms. Some will go to all meetings and/or some will adopt all new behaviors—others will not. Some will be in the project from its inception and others only for a few months. Some farmers in the treatment group will take no active part at all.

In Kenya, the impact study originally defined participation as residence in one of the areas in which the projects intended to work. As it turned out, there was much variability in the degree of farmers' exposure to project-supported interventions, such as participation in growers' groups, use of inputs or services, participation in contract sales, and the development of new relationships and levels of trust. Because this range of difference was not incorporated into the definition of participation, the understanding of the quality of project (in so far as it could be shown) was lessened.

COMMON BUT SERIOUS PROBLEMS

Apart from the five inherent challenges discussed above, which are particularly likely to occur in the study of EG projects, numerous other problems can affect many types of impact assessments. These common problems have been covered elsewhere[8] and a complete discussion is beyond the scope of this paper. However, three types of problems deserve mention here because they are particularly widespread and can negatively affect the usefulness and credibility of an impact assessment.

---

[7] Rees Warne, "BDS Report for Azerbaijan IGP-BDS Project # 1064; Mercy Corps Cluster Access to Business Services (CABS) Program in Rural Azerbaijan," October 2007 (Micro Report # 83).

[8] Madhabi Chatterjee, "Grades of Evidence; Variability in Quality of Findings in Effectiveness Studies of Complex Field Interventions," *American Journal of Evaluation*, (Sept 2007) See also Jim Tanburn, *The 2008 Reader on Private Sector Development*, ITC 2008 and Lucy Creevey, "Common Problems in Impact Assessment Research," IAPS # 7.

The first of these common problems relates to **timing**—when to conduct both the baseline survey and the follow-up survey. Ideally, assessment research should be designed before the project enters the field and baseline data should be collected before participants have been affected by the intervention. However, this often is impossible. In fact, given the nature of EG projects, the ideal of pre-intervention baseline scheduling is regularly contravened since rapid evolution of the project in the first years leads to situations in which initial baseline samples do not include individuals who benefit from the mature group of interventions.

Another aspect of the timing problem is when to schedule the follow-up survey. The convention is to wait until two years after the baseline survey. A shorter time lapse between surveys is likely to be too brief to allow many of the program's impacts to be captured. Yet some projects are simply too short to allow impacts of project interventions to emerge fully within the timeframe of the program. In the case of Brazil, the evaluability assessment team concluded that expending resources on a full impact assessment could not be justified, primarily because the project length was too short—only two years.

The Kenya and Zambia studies also suggest that two years may be too brief an interval to measure the program's full impact, since all forms of project impact may not yet have emerged and become measureable. Although project planners and donors wish to know as early as possible if the project interventions were successful so that future projects may be planned, a longer time lapse provides a large enough window for impacts to appear. In addition, it may be more interesting to see how participants fare once the project personnel are gone. A three-year interval between surveys or a second follow up survey a year or two later are possible remedies for this problem, but by that time it may be getting difficult to locate the original respondents. Clearly there is no definitive solution to the timing problem; trade-offs must be weighed in each case.

The second common issue undermining research is the occurrence of **exogenous shocks** that seriously alter the project environment. In some projects, sudden changes in the external environment—related to weather, markets, social conflicts, or politics—may lead to shifts in intervention strategy. This was the case in Zambia, where changing cotton prices and excess rain brought about shifts in project emphases and locations. The case in Azerbaijan is even more telling. In the year before the follow up study, the project regions were devastated by anthrax and hoof-and-mouth disease. In addition, a widespread avian flu scare affected both the consumption of poultry products and the health of poultry flocks. The result was a sharp drop in farmer income, which was completely counter to desired project outcomes. While it can be argued that even in adverse circumstances, program participants should do better than non-participants, some types of exogenous shocks are so extreme and unexpected that they nullify the basic assumptions underlying the project design.

Finally, a number of familiar problems relate to **data quality**. Serious problems with data quality can occur at many possible points, including during the design and testing of data collection instruments (for example, with questionnaires), selection and supervision of field survey teams, data entry and cleaning, and data analysis. Unfortunately, researchers may not become aware of data quality problems until it is too late to go back and correct them. This can make it impossible to answer some of the key evaluation questions. It can also compromise the credibility of the findings. Data quality should be carefully monitored in all types of impact assessment, including assessments of EG projects.

A FRAMEWORK FOR MOVING FORWARD

An ideal framework for evaluating the effectiveness of new-generation EG projects produces information that is both useful and credible. The evaluation results are useful in the sense that they answer meaningful questions and provide information that is needed for strategic decision making about future programming. The information is credible in the sense that it does not point toward false conclusions, but instead provides reliable evidence about the extent to which a project achieved its goals.

Given the challenges detailed in the previous section, the best approach for striking a balance between usefulness and credibility is to assemble a portfolio of evaluation methods. Such a portfolio of methods should include both standard impact assessment methods to answer micro-level questions about impact based on well-defined counterfactuals, as well as alternative methods to provide credible answers to important questions based on a preponderance of evidence. This **effectiveness evaluation** approach may be defined as measuring the extent to which targets are being met and detecting the factors that hinder or facilitate their realization. It also involves establishing cause-effect relationships about the extent to which a particular policy (or set of policies) may be shown to produce the desired outcome.

BALANCING USEFULNESS AND CREDIBILITY

Usefulness comes from having the questions guide the study. A useful evaluation informs and improves decision making by building credible evidence to answer important questions. The first step in any evaluation is to identify the questions that need to be answered. The questions should be the motor that drives the selection of evaluation methods. This stands in sharp contrast to an approach that begins with the selection of a specific evaluation method then limits the scope of the evaluation to the kinds of questions that can be answered using that method. For the results of an evaluation to be useful, the questions should determine the methods, not the other way around.

The questions to be addressed in an **effectiveness evaluation** will depend on the interests of the decision maker(s) requesting the study. Donors will likely want to know about achievements in broad program areas, while project implementers may be more interested in determining the relative effectiveness of alternative delivery mechanisms. Even so, there are generally two categories of questions:

- Change-related questions: To what extent were project goals achieved?
- Attribution-related questions: What is the evidence that the project led to those changes?

Standard impact assessment methods based on experimental and quasi-experimental statistical designs are good for answering attribution-related questions, especially those posed at the microeconomic level of individuals, households, or firms. As discussed below, internal validity is a major strength of these standard impact assessment methods. However, when the project includes policy-level interventions, extensive spillovers, and/or ambiguities in participant identification, then certain impact questions may be difficult or impossible to address with standard impact assessment methods.

Important questions about the extent to which a project achieved its overall goals should not be ignored just because it is not possible to use standard impact assessment methods to build the case for attribution. No matter what questions are being asked, valid answers require careful selection of appropriate and rigorous evaluation methods. When appropriate methods are used to build credible evidence, then evaluation can be a valuable tool for improving development programming and aid effectiveness.

DEGREES OF EVIDENCE PRINCIPLES

The degrees of evidence framework provides guidance for selecting alternative ways of achieving useful results, based on the belief that no one approach is ideal in all situations. All evaluations can be criticized, but carefully selecting methods can increase the credibility of the results. Evaluative rigor and credibility can be strengthened by adhering to the following principles:

- The evaluation should be grounded in a plausible causal model.
- The evaluation methods should be assessed relative to four standards of methodological validity.
- The evaluation findings should be triangulated to ascertain the preponderance of evidence.
- The evaluation methods should follow sound data collection practices.
- The evaluation methods used, with their strengths and weaknesses, should be made transparent to the end user.

The degrees of evidence principles provide an approach for understanding methodological rigor, which is helpful both for understanding the relevant distinctions between evaluation methods and for matching evaluation methods to the questions asked, the level of credibility required, and the resources made available.

CAUSAL MODELING

A causal model is a logical chain of events connecting the project activities to the final impacts the project is designed to achieve.[9] It consists of a series of cause-and-effect relationships starting with the program activities and ending with intended impacts. Along the way, the causal chain passes through a series of intermediate linkages, including the project's expected outputs and its short-, medium- and long-run outcomes. The project's impacts can be defined as the subset of project outcomes that can be said, with a reasonable degree of confidence, to be attributable to project activities.

The relationships in the causal model reflect the impact hypotheses to be tested as well as the underlying assumptions about how impacts occur. Not only does the causal model provide the conceptual framework for designing the evaluation, but it also undergirds the ability to assert causality in the findings and attribute the observed changes to the project. As learning accumulates over time, the causal model underlying an entire program area becomes better understood.

In most cases, the detailed causal model is developed during the earliest stage of the impact assessment, under a process known as an evaluability assessment.[10] Given the evolving nature of new-generation EG

[9] Woller and Downing, *op. cit.*

[10] Elizabeth Dunn, "Planning for Cost Effective Evaluation with Evaluability Assessment," Impact Assessment Primer Series #6, January 2008. See also Joseph S Wholey, "Evaluability Assessment," in Joseph S Wholey, Harry P. Hatry III and Kathryn E Newcomer, eds., *Handbook of Practical Program Evaluation*, Second Edition, John Wiley, 2004, pp. 33-62.

projects, the causal model must be periodically revisited and adjusted to reflect changes in project activities. Care must be taken to design the baseline study in such a way that it retains its relevance over the life of the impact assessment.

METHODOLOGICAL VALIDITY

Methodological validity refers to the truthfulness of the claims that are made about impact. Given that standard impact assessment methods cannot be used to evaluate every important component of an EG project, a clear set of criteria is needed to assess the credibility of alternative methods and the results they generate. Methodological validity can be assessed along four distinct dimensions: internal validity, external validity, construct validity, and statistical conclusion validity.

Internal validity refers to whether inferences concerning causal relationships can be said to be true. It is the extent to which the findings of an evaluation accurately represent the causal relationship between an intervention and an outcome or impact, given the particular circumstances of the project.

Internal validity is an essential criterion for attribution-related impact questions. To establish a strong case for internal validity, an impact study must meet three conditions:

- There is a plausible theoretical model underpinning the proposed cause-and-effect relationships.
- Alternative explanations for the observed change, such as selection bias, can be ruled out.
- There is an established counterfactual that allows observed changes to be attributed to the project. All impact assessments are based on an explicit or implicit comparison to the counterfactual; that is, they ask the question, “What would have happened to the treatment group in the absence of the treatment?”

Thus, internal validity requires that observed changes can be attributed to the program and not to other possible causes. Alternative explanations must be ruled out. The evaluation must establish a valid counterfactual. Since the counterfactual is always unobservable, estimating it is the core challenge of impact assessment.

Because experimental and quasi-experimental methods create a clear and credible counterfactual, they are the preferred methods for establishing internal validity. Moreover, experimental methods are superior to quasi-experimental methods in eliminating selection bias, since they rule out this alternative and competing explanation for a finding that project participants have better outcomes than nonparticipants.

As discussed earlier, however, some important questions cannot be answered using experimental or quasi-experimental methods. When evaluation questions are at the level of a market, community, or country, there will be few or no plausible counterfactuals. Yet questions about impact at these levels should not be ignored. Instead, they should be approached by triangulating alternative results in an effort to establish a preponderance of evidence for impact.

External validity measures the extent to which the findings obtained from an evaluation conducted under particular circumstances can be generalized to other contexts and settings. If the findings from a particular evaluation do not apply to other situations, then the findings are said to lack external validity in those other situations.

External validity is often a matter of degree. It requires judgment about how well different contexts match each other. The problem with some statistical methods is that they might narrow the questions being asked so sharply that the answers become unique to the specific project and setting.

Construct validity is the degree to which legitimate inferences can be made from the evaluation study to the underlying theoretical concepts (variables) included in the causal model. When assessing construct validity, researchers ask, "Were the treatment, outcome and other mediating variables appropriately defined and measured in the empirical work?"

The cause (treatment) and effect (outcome) variables correspond to the underlying theoretical concepts, as expressed in the causal model. If the treatment, outcome or other variables were not appropriately defined or measured in empirical work, then inferences about the underlying theoretical model cannot reasonably be made.

Statistical conclusion validity assesses whether the appropriate statistical tests were applied and reported accurately. It does not relate to the results, just to the process of getting to the results. The results may not be significant, but they can still have statistical conclusion validity. The researchers must have correctly applied statistical methods and identified the statistical strength and confidence level of their results. In other words, the degree of uncertainty about the conclusions must have been correctly identified.

TRIANGULATION OF EVIDENCE

Conclusions about impact become stronger when they are supported by multiple sources of evidence. Mixed method evaluation designs using different combinations of quantitative and qualitative methods allow researchers to triangulate toward more credible evaluation findings.

Combining surveys with one or more qualitative methodologies can yield credible results and provide richer insights while saving time and cost. Using multiple methods of data collection such as surveys, focus group discussions, and in-depth interviews allows for cross checking of findings to validate data and ensure the reliability of responses.

In general, the use of mixed methodologies is the preferred approach to impact assessment of EG programs, as it is the approach most likely to yield credible in-depth findings. Triangulation can be used to create a preponderance of evidence about whether the project achieved (or did not achieve) its goals.

SOUND DATA COLLECTION PRACTICES

All evaluation methods should be implemented following accepted good practice in data collection, including the use of competent researchers and the implementation of sound quality control measures. Impact assessment surveys for EG projects face the common challenges of data collection cited in a previous section and can be easily undermined by difficulties experienced in the collection of valid data. Important measures for overcoming these difficulties include:

- Use of a feasible and well-designed research protocol (related to construct validity);
- Valid sampling procedures (related to statistical conclusion validity);
- Strong, well-tested survey instruments; and
- Experience, training, and close supervision of field personnel.

METHODOLOGICAL TRANSPARENCY

Methodological transparency is essential. Evaluator must ensure that methods are well documented and their weaknesses and related implications are identified, including:

- Anticipating and addressing threats to validity in the design phase;
- Following a well-designed research protocol in the implementation phase; and
- Describing the methods used and exposing threats to validity in the dissemination phase.

One characteristic of rigor is the extent to which the weaknesses of the study are acknowledged and explained by the evaluators. The end user of the evaluation findings needs to understand the design choices that were made and trade-offs involved in these choices.

DESIGNING EG PROJECT EVALUATIONS

This section illustrates how the degrees of evidence principles can be applied to design EG project evaluations. The principles are applied to the case of a hypothetical EG project to benefit small-scale mango farmers in an East African country. This example is typical of recent USAID EG projects, in that it promotes economic growth with poverty reduction through a set of value chain interventions. To evaluate the project, standard impact assessment methods based on attribution are combined with alternative evaluation methods to ascertain the preponderance of evidence, resulting in an evaluation of project effectiveness that is both useful and credible.

EXAMPLE OF A TYPICAL EG PROJECT

This typical (but hypothetical) project promotes national economic growth and the reduction of rural poverty by intervening to strengthen an East African country's mango value chain and increase smallholder participation in that value chain. The project thus seeks both to make the country more competitive in domestic, regional, and international markets for mangos and mango products and to help its smallholders reap higher rewards from their mango production. To do so, the project works with companies that export fresh mangos and other mango products to Europe. It also interacts with local and regional mango processors and supermarket chains. In addition, it seeks to strengthen farmers' associations by working through local NGOs. The project uses a value chain approach to link small-scale farmers to more remunerative markets, thereby raising their production, productivity, sales, and product quality.

New-generation EG projects differ from the older style of projects. Older projects often worked directly with small-scale producers, trying to improve their production methods and spread the use of improved technologies. Newer projects are designed to benefit small-scale farmers indirectly by working through private sector actors and emphasizing improvements in small-scale producers' market integration by strengthening their vertical and horizontal market linkages.

The anchor for the hypothetical mango project is a lead firm that exports mangos and mango products from East Africa to Europe and the Middle East. For the success of its business, the lead firm requires timely delivery of a given quantity of mangos of a certain variety that meet specified quality standards. The value chain has had problems associated with the types of mangos grown, low productivity, plant diseases, the collection of adequate quantities of produce in one place at an appropriate time, and transportation of fruit from the farms to packaging and processing facilities. To deal with these problems, the lead firm is encouraged to enter into contracts with groups of mango growers to deliver the right quantities and qualities of fruit according to an agreed-upon schedule. The firm will guarantee the farmers a certain price for their mangos and will provide certain agricultural inputs and services (directly or through agents) to help the farmers meet their contractual obligations.

To facilitate their arrangements with the lead firm, the small-scale farmers are encouraged to group themselves into farmers' associations that help disseminate production and marketing information and assist in product aggregation and delivery to the lead firm. The role of the EG project is to help the market

work better by demonstrating the potential benefits to all concerned of improving the workings of the value chain. In the long run, the improved value chain should be sustainable and operate without further intervention.

The success of this kind of this new-generation EG project will depend in large part on how well it is designed and implemented, in particular on whether it succeeds in developing trust among the various participants and ensuring that benefits accrue to all those involved. The success of the project also depends on the business environment in which the mango value chain operates at the local, sectoral, national, and international levels. The applicable laws and regulations at all these levels must enable the project to meet its goals. Some EG projects directly address these policy and business environment issues. In the case of the mango project, project implementers are working with government agencies and business leaders to end the monopoly on imports of fertilizer and other agricultural inputs granted some years ago to a parastatal concern that raises the prices that farmers must pay for such inputs. They are also involved in efforts to help mango farmers comply with new and higher sanitary and phytosanitary standards imposed by European importers fruit imports.

A successful EG project such as this hypothetical example of mangos in East Africa will raise the income and well-being of household members of participating small-scale producers. It will also facilitate the development of the value chain as a whole, permitting larger quantities of better quality products to be sold at higher prices in more distant markets. One important job of effectiveness evaluation is to determine whether the project succeeded in doing this. An equally important task is to determine why the project succeeded or failed to meet this objective.

UNDERSTANDING EVALUATION RESULTS

The most important issue to be explored via impact assessment is what the project achieved at the level of producers and households. The income and welfare of participating households should improve and the improvements should be traceable, at least in part, to the activities of the project. To determine whether this is the case, standard impact assessment methods (experimental or quasi-experimental analyses) that employ a counterfactual should be used, despite the many issues they raise, since they are the only methods that can create a strong case for attribution.

At the same time, new-generation EG projects also try to promote economic growth and improve the workings of the value chain as a whole. Standard impact assessment methods cannot tell us whether this kind of objective was achieved because a credible counterfactual is not available.[11] Conclusions about project impacts at this level must instead rely on ascertaining the preponderance of evidence that the desired changes occurred and could be credibly linked to the project interventions.

Standard impact assessment methods indicate whether impacts occurred at the level of the ultimate beneficiaries targeted by the project. If impacts are found with a credible level of certainty, that must mean, in general, that the project succeeded because it was well designed and implemented in a business environment that supported its success. However, more precise information on the reasons for project

[11] If the project aims to promote the export of mangos from an East African country, one could conceivably compare the development of the East Africa mango value chain with that of the East African avocado value chain or the mango value chain in West Africa. Something could be learned from such a comparison, but the many differences across products and countries would make causal inferences risky.

success can add significant value to the impact findings by shedding light on how and why impacts occurred.

If no impact is found, there are, broadly speaking, three possible explanations:

- The project had significant flaws in its design and/or implementation.
- The project was well designed and implemented, but negative factors in the business environment or "acts of god" such as natural disasters kept it from succeeding.
- The project did succeed but the impact assessment study failed to identify its success.[12]

When the findings of an impact evaluation indicate that the project has no impact, all three of these possibilities should be considered.

RECOMMENDED EVALUATION PROCEDURES

In the light of these considerations, which grow out of our group's experience in carrying out impact assessments of USAID's new-generation EG projects, we recommend that the following procedure be followed in future impact assessment studies:

1. Undertake an evaluability assessment[13] early in the project's life, in which you clarify, with project personnel, the initial causal model for the project (involving goals, activities and expected outputs, and outcomes and impacts).
2. Develop a research plan that includes a causal model, hypotheses to be tested with their theoretical underpinnings, "domains of analysis" clarifying what to measure (indicators) and how, and what outcomes and impacts are expected to emerge between the baseline and endline surveys. The research plan should set out the design of the evaluation activities, including the two surveys, qualitative methods to be used, the timing of these activities, and other aspects of evaluation design.
3. Design a baseline survey, using experimental methods if possible (this requires the cooperation of the project implementers and local parties to permit participants to be chosen at random) or quasi-experimental methods if not. The survey should collect information on a sample of enterprises and households in an area where the project expects to work, as well as on a comparable sample in an area in which no project activity is planned.[14]
4. Carry out the baseline survey in cooperation with a local firm. Evaluators must closely supervise data collection to ensure their high quality. The sample size should be larger than will be needed for the final analysis to allow for sample attrition.

---

[12] One reason for this could be that members of the control group received spillover benefits that improved their performance relative to that of members of the control group. Such an event indicates success for the project, but it undermines the classic comparative assessment methodology.

[13] See Elizabeth Dunn, op. cit. "An evaluability assessment can save valuable time and resources by determining whether it makes sense to conduct an impact assessment and, if it does make sense, by gathering the preliminary information needed for planning a well targeted and cost-effective effort. A site visit is usually required because much of the work involves extensive interviews with program stakeholders including the sponsoring organization, headquarters and on-site program staff, program partners, and intended program beneficiaries. In general, it can take one to three weeks for a team of two people to complete a routine evaluability assessment. This investment will more than pay for itself by leading to impact assessment that are valid, efficient, and useful."

[14] The control group—those not involved in the project—may be in the area where the project operates, but in this case, the likelihood of spillovers is much greater.

5. Analyze the results of the baseline survey.

6. To complement the baseline survey, carry out qualitative research among the farmers/artisans/microentrepreneurs, traders, processors, marketers, and more. This research can take the form of interviews, group discussions, and in-depth histories.

7. In the interim period, before you conduct the follow-up survey, stay in touch with the project through process evaluation. It is important to take note of and understand changes that occur in project strategies, activities, partners, and areas of operation.

8. After two or three years (not sooner), plan the follow-up survey and complementary qualitative research. If the project has continued to work in the treatment areas selected for the baseline survey, these areas should be resurveyed in the follow-up survey. Supervise the follow-up survey closely to ensure that high-quality data are collected. Since a set of panel data is desired, data collected from particular respondents during the two survey rounds must be consistent with each other.

9. If, however, project activity has ceased or shifted to other areas, no follow-up survey is warranted. In this case, focus on the reasons for the changes (such as the extent to which they reflect failures of earlier approaches or learning that results in improvements) and on what is being achieved in the areas to which activity has shifted. Since no baseline survey will exist for the new areas, the study will need to rely on qualitative evidence and can produce only conclusions based on the preponderance of evidence.

10. Write up a final report based on all of the above.

The degrees of evidence framework tells us what constitutes a truly rigorous impact assessment of a new-generation EG project such as our hypothetical East African mango example. To know whether the mango-growing smallholders benefited from the project, either an experimental or a quasi-experimental study must be conducted at the level of their mango-growing enterprises and households. While the experimental approach is likely to have greater internal validity, applying that approach in studying a USAID EG project is often impractical and/or politically or ethically unacceptable. In that case, the quasi-experimental approach should be used, with great care given to minimizing its well-known shortcomings. In addition, qualitative research at the producer level is useful for deepening understanding of the quantitative findings.

Whatever approach is taken at the firm/household level, additional study will be needed both to determine the reasons that particular types of impact were or were not realized, and to understand what impact the project may have had at the level of the value chain. Without such additional study, the impact assessment, however strong it may be in terms of internal validity, will be weak in external validity. It will be difficult to draw legitimate inferences from a study conducted in particular circumstances and apply them to other times and places. Even though study of the project's achievements at higher levels—such as its impact on the value chain as a whole—cannot produce definite proof of attribution but must rely on qualitative methods and can only yield a preponderance of evidence to suggest (but not demonstrate) attribution, this kind of investigation is necessary to evaluate the project as a whole. Moreover, mixed method evaluation designs using combinations of quantitative and qualitative methods can produce a triangulation of findings that confirm (or cast doubt on) the findings from a single method.

CONCLUSION

The experiences of USAID's PSD IAI offer important lessons for how to improve the effectiveness and accuracy of impact measurement for EG projects. This body of work shows that effectiveness research must be multifaceted, flexible, comprehensive, and ongoing throughout the life of the project. These conclusions do not imply that earlier discussions of how do to evaluations or how to conduct an impact assessment reach invalid conclusions. Many of these discussions (some cited as references in this paper) did make clear important considerations in doing research and pointed out common mistakes that those who have tried to do this research have incorporated into their work. Earlier studies also stressed the usefulness of multiple methods to capture the complexities of EG projects. But the carefully defined and controlled impact assessment studies done under this initiative and reviewed here demonstrate the ways in which new-generation EG projects require interlocking pieces of research with methods differing depending on the particular project goal considered. Using the degrees of evidence approach, researchers can assess what each piece of research will be able to show and the extent to which it can demonstrate impact or merely indicate a project outcome that is tied to the project interventions by the preponderance of evidence.

APPENDIX 1: PSD-IAI IMPACT ASSESSMENT CASE STUDIES

USAID's Kenya Business Development Services (KBDS) and Kenya Horticulture Development (KHDP) projects. Planning for the study began with an evaluability assessment in January 2003. Fieldwork was carried out from November 2004 to January 2005 and again from December 2006 to February 2007. Data entry and analysis followed, and a final report was completed in December 2007. This was a quasi-experimental study with quantitative (sample survey) and qualitative (interview and focus group discussion) components that studied both program participants and a control group in an attempt to determine whether project activities involving smallholder cultivation of avocados, mangos, and passion fruit had positive impacts on the value chains for smallholder tree crops, on the enterprises of participating smallholders, and on participants' households.

USAID's Productivity, Finance and Improved Technology (PROFIT) project in Zambia. The study began with an evaluability assessment in November 2005. Fieldwork was carried out from August to November 2006 and again from November 2008 to April 2009. The Zambia study was quasi-experimental in design and included qualitative and quantitative components. The study tried to determine whether project activities involving smallholder cultivation of cotton, smallholder beef cattle herding, and retail distribution of farm inputs to smallholders had positive impacts on the cotton and beef value chains, on farm input distribution, and on the welfare of the participating farmers and their households.

USAID's Micro and Small Enterprise Trade-Led Growth Project (CRESCE) in Brazil. This study began with an evaluability assessment and consisted of semi-structured individual interviews and focus group discussions applied to a relatively small sample conducted first in 2006, with a second round in 2007. The study explored the affects of brokering commercial linkages, upgrading, facilitating access to financial services, and identifying barriers to sector competitiveness on the business and net business income of participants in three value chains: beachwear in Salvador, Bahia; cashew nuts in Barreira, Ceará; and honey in Simplício Mendes, Piauí.

USAID's Cluster Access to Business Services (CABS) Azerbaijan. This study was quasi-experimental in design and included qualitative and quantitative components. The baseline survey was carried out in 2003 and the follow up in 2006. The study assessed the impacts of organizing farmers into small clusters through which they could more effectively receive information on improving the care and marketing of their animals/poultry, and training veterinarians in providing more and better services to the farmers. Intended results included more purchases of (more diverse) veterinary services by farmers, higher quality animal products and increased net profits for both farmers and veterinarians.

USAID's Growth-Oriented Micro Enterprise Development Program (GMED) in India. The evaluability assessment for this study was conducted in 2005. The study design included a quasi-experimental impact assessment, a process evaluation, and a qualitative component that included individual case studies and focus groups. The baseline survey was conducted in early 2007, followed by a second round in early 2009. Initially, the evaluation covered project impacts in three value chains. Because of findings from the process evaluation, the second round survey and qualitative data collection focused on one value chain: smallholder production of fresh vegetables for domestic supermarket chains.