



USAID
FROM THE AMERICAN PEOPLE

COMMON PROBLEMS IN IMPACT ASSESSMENT RESEARCH

IMPACT ASSESSMENT PRIMER SERIES
PUBLICATION # 7

PRIVATE SECTOR DEVELOPMENT IMPACT ASSESSMENT INITIATIVE

JUNE 2008

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

COMMON PROBLEMS IN IMPACT ASSESSMENT RESEARCH

IMPACT ASSESSMENT PRIMER SERIES PUBLICATION # 7

PRIVATE SECTOR DEVELOPMENT IMPACT ASSESSMENT INITIATIVE

DISCLAIMER

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

- COMMON PROBLEMS IN IMPACT ASSESSMENT RESEARCH..... 1**
- Introduction1
- Timing.....1
- Spillover Effects2
- Selection Bias.....3
- Ensuring Good Performance by the Local Research Partner5
- Unanticipated External Events.....6
- Conclusion6

COMMON PROBLEMS IN IMPACT ASSESSMENT RESEARCH

Introduction

Doing an impact assessment of a private sector development (PSD) program is inherently challenging. Doing it the “right” way—such that it satisfies minimally acceptable methodological standards—is more challenging yet. During the course of planning and implementing an impact assessment, it is not uncommon for researchers to confront any number of problems that have serious implications for impact assessment methodology and, consequently, the validity of its findings.

This contribution to the Impact Assessment Primer Series draws on selected case studies to describe common problems researchers confront in doing impact assessments of PSD programs, how researchers have dealt with these problems, and the related implications for the validity of the assessment findings. The frame of reference for discussing these issues is the minimally acceptable methodological standards established by the Private Sector Development Impact Assessment Initiative (PSD-IAI) in earlier Primer Series papers.¹ The question this paper asks is whether the methodological approaches taken in the selected case studies satisfied minimally acceptable standards for a rigorous impact assessment.

The impact assessment problems discussed in this paper include: timing, spillover effects, selection bias, capability of local research partners, and unanticipated external factors, such as climatic disasters. These do not capture all the problems impact researchers face, but they do capture some of the more common ones.

Timing

The timing of the impact assessment may seriously affect the validity of its findings. Ideally, a set aside for an impact assessment is incorporated into the original program budget that includes funding for a technical expert to set up the impact assessment early in the program cycle. More commonly, however, the decision to do an impact assessment occurs after the program is already underway. This can cause a number of problems. To begin with, the baseline may come too late to capture impacts that have already occurred, resulting in an understatement of actual program impacts. The longer the time lag between program launch and the baseline research, the greater the probability that the impact assessment fails to capture certain program impacts.

Even more striking examples of the problems resulting from delaying the start of research are provided by cases in which the impact assessment is done either near the end or after the end of a program. In these cases, there is no possibility of doing a baseline study, or, indeed, of getting any longitudinal data. Everything depends on a one-time set of research activities and often entails a heavy reliance on retrospective questions.

One example of this is a recently completed impact analysis of the three-year AT India project in Uttarakhand. The project’s objective was to create a business development service market among micro dairy farmers so as to expand their businesses and increase their net profits. Because of serious flaws in the baseline research (see the section below on Selection Bias), the baseline data had to be abandoned. The final report, therefore, consisted entirely of retrospective questions taken from a post-project survey of farmers, BDS providers, and dairy product consumers.²

¹ See Don Snodgrass, (2006), “Assessing the Impact of New Generation Private Sector Development Programs,” Impact Assessment Primer Series Publication #1, http://www.microlinks.org/ev_en.php?ID=15094_201&ID2=DO_TOPIC. See also Lucy Creevey and Gary Woller, (2006), “Methodological Issues in Conducting Impact Assessment Research, Impact Assessment Primer Series Publication #2 http://www.microlinks.org/ev_en.php?ID=15095_201&ID2=DO_TOPIC. See also www.microlinks.org/psdimpact.

² Retrospective questions ask respondents to recall events in the past.

Although the researchers were able to gather data on a participant and control group of farmers and BDS providers, the results were not as complete or credible as they would have been had the baseline been done correctly.³

There is nothing wrong, per se, with retrospective questions. If used judiciously, they can be a valuable research tool, particularly in cases where a baseline either was not done or could not be used. Because retrospective questions rely on recall, however, they tend to involve a greater amount of measurement error than contemporary questions. The amount of measurement error is greater the longer the recall period and the more detailed the information requested. Retrospective studies, therefore, are expected to be far less accurate than studies including a baseline and follow-up.

In terms of cost, end-of-program assessments are considerably less expensive. With no risk of panel attrition (loss of sample participants due to relocation, death, inability to locate, etc.), sample sizes need not be as large, nor are the costs for local research partners or foreign technical experts anywhere near as great. End-of-program evaluations can be useful if done carefully and appropriately limited in scope, but they lack the depth, accuracy, and statistical validity of a well-done longitudinal assessment.

Within a longitudinal assessment, both timing for the baseline (discussed above) and timing of the follow-up research are crucial. Of course, one option is to conduct the follow-up research near or soon after program wrap-up. This approach permits an assessment of program outcomes and impacts during the life of the program.

The follow-up research, however, need not necessarily wait until the program concludes. For long programs (e.g., those lasting three or more years), it may not be feasible to wait until the program concludes. Donors or program management may not be willing to wait four, five, or more years to see results from the impact assessment. Program management may also wish to use the assessment findings to make mid-stream adjustments in the program design or implementation. There is also the risk that too long a timeframe will cause loss of interest (out of sight, out of mind), particularly if the donor agency or program management experiences a change in management.⁴

If program sustainability is an important concern, one might consider delaying the follow-up research for a period of time after the program has concluded. Findings of impact at or near program conclusion do not necessarily mean that the impact will be sustained over time. Experience shows that program beneficiaries are usually very pleased when the program is just starting, underway, or just finished. They are still reacting to having resources and attention accessible to them that they did not have before, and they still have the energy, enthusiasm and hope with which they embarked on the new venture. The only way to determine whether the benefits experienced by the program clients are sustained over time is to carry out the follow-up research at some point after the program has concluded.⁵

Spillover Effects

A second common problem occurs when program benefits spill over to non-program participants. An example is the recently completed impact assessment of the Cluster Access to Business Services (CABS) program in Azerbaijan, which seeks to “improve profitability for clusters of rural poor and women micro-entrepreneurs by increasing access to a network of trained veterinary and production advice service providers” The baseline study, conducted a year

³ Lucy Creevey and Momar N’Diaye, (2007), “Development of a BDS Market in Rural Himalaya: A Research Report,” forthcoming.

⁴ The PSD-IAI balances the need to allow sufficient time for impacts to emerge with the need to produce results by opting for a two or three year timeframe.

⁵ If not possible to do follow-up research after the program has concluded, a second best alternative is to evaluate whether project participants have adopted behaviors that are considered preconditions for sustainability. In an impact assessment of tree fruit producers in Kenya, for example, PSD-IAI researchers evaluated which producers have planted new fruit trees and which have shifted from traditional crop varieties to varieties sought by exporters.

after program launch, showed significantly higher net profits for the treatment group veterinarians, but this difference had disappeared by the time the follow up study took place.⁶

A couple of factors explain why the difference between treatment and control veterinarians had disappeared in the interim between the baseline and follow-up. During the interim, client veterinarians had shared what they learned in their training sessions with their colleagues who were not being trained. It also turned out that during the baseline period, client veterinarians were the only ones with access to high quality animal drugs, which they obtained from the program. During the interim, however, all the veterinarians in the study gained access to the same high quality drugs. While such spillover effects are generally desirable from a program's perspective, their downside is that they produce a contamination in control group, which makes it look as if the program has had less impact than it actually did.

The researcher heading this impact assessment incorporated all these factors in her analysis and included retrospective questions in her baseline survey to tap this information. Nonetheless, the existence of spillover effects almost certainly resulted in systematically lower impact findings for the program.

In the case of the AT India project, researchers found that the critical variable distinguishing dairy farmers who had improved their sales and incomes was a key strategy of the project—forming a Self Help Group (SHG), which assisted the farmers in collecting and marketing the milk products. Subsistence dairy farmers who belonged to an SHG had significantly higher net profits than those who did not belong. Participants in the AT India project, moreover, were significantly more likely to join an SHG.

What the findings could not clearly disclose was whether control group farmers who formed SHGs did so because they had heard that treatment farmers in neighboring villages (in the project) had enjoyed good success with SHGs. Focus group discussions indicated that the village women did hear about women in other villages through communication networks connecting the villages. The situation was even more complicated than this. While SHGs had existed in the region before the project started, it appears that many more were created as a result of the project. The project also facilitated the previously existing SHGs in assuming new marketing roles and provided them with access to business development services (BDS), such as vaccine camps for their animals and credit.

Project participants in the SHGs were more likely to buy seeds and vaccines than non-participants in SHGs, but there was not a significant difference overall in net profits between treatment and control SHG members. The possibility that there was a spillover effect, therefore, appears quite high. Researchers who had been forced to rely on retrospective data were unable to tease out how, when and why women in the control group had formed SHGs and whether those SHGs that had existed pre-project had changed what they did as a result of their neighbors' activities.

The study's findings, therefore, could not state authoritatively what the project's impact had been aside from endorsing the effectiveness of SHG formation and the empowerment these groups through assistance in setting up collection and marketing schemes and access to BDS. The project's overall impact on the micro dairy industry was likely in greater than could be credibly demonstrated.⁷

Selection Bias

One of the greatest challenges in doing a high quality impact assessment is identifying statistically valid treatment and control groups. The best method of group selection is the experimental method in which membership in the

⁶ See Rees Warne, "Cluster Access to Business Development Services in Rural Azerbaijan: Final Report," (2007), forthcoming.

⁷ Lucy Creevey and Momar N'Diaye, op.cit.

treatment and control groups is determined via random assignment. Where experimental methods are not feasible, quasi-experimental methods are a second-best alternative.⁸

In both experimental and quasi experimental research, the importance of getting group selection right cannot be overemphasized. The cost of not getting it right is introducing selection bias into the sample. Selection bias occurs when control group members systematically differ from treatment group members in terms of either observable or unobservable characteristics.⁹ In most PSD programs, participants either select themselves by coming forward voluntarily or are specifically recruited by program administrators because they represent the most suitable participants. These procedures facilitate program implementation and contribute to good results, but they make it difficult to tell if observed differences in impact are the result of the program intervention or the specific characteristics of program beneficiaries.

In the case of the AT India program discussed above, the program targeted poor female micro dairy entrepreneurs from two specific watershed areas. In selecting the control group, however, researchers chose a random sample of all dairy farmers from nearby areas. The treatment and control groups, therefore, differed significantly in terms of characteristics such as gender, socio-economic status, and access to inputs, thereby confounding any attempt to make comparisons between the two groups. As a result, the baseline had to be discarded. While this might appear an extreme example of selection bias distortion, it is less unusual than one might presume. Selection bias ranging from minor to severe is the rule rather than the exception in doing impact assessment.

Selection bias can even find ways to creep into experimental research designs. Consider the case of an experimental impact assessment of a microfinance program in South Africa. The assessment examined the impact on institutional performance and client well-being from relaxing the program's lending criteria and granting loans to a randomly selected group of marginal applicants who would have been denied loans using the normal lending criteria. (Marginal applicants were randomly assigned to receive or not receive loans.) During implementation, however, loan officers rejected several of the applicants randomly assigned to the treatment group. It turns out that their bonuses were based on client repayment rates, and without some kind of guarantee of hold-harmless, they rejected approximately 47% of all the treatment group members who they deemed not to be credit worthy.¹⁰

The best way to avoid selection bias is through careful research design and close supervision when the research team implements the sampling criteria. Failure to control for selection bias produces systematically biased and potentially unreliable results. Among the many possible mistakes in impact assessment research, getting the sample wrong may be one of the most serious. It makes little sense to invest the time and money in doing an impact assessment if it produces unreliably biased findings.

When a sampling error is discovered, it may be possible to correct for it in subsequent research steps or in the analysis and presentation of final results. In all cases where it occurs, researchers should disclose the source, degree (to the extent it is knowable), and the implications of the selection bias.

In the case of the AT India program, researchers chose to establish a new sampling procedure for the follow-up survey. They revised the survey to include questions that looked retrospectively at the dairy businesses, access to BDS, membership in women's village groups, and other important items at the outset of the project and in the most recent period before the survey. This strategy allowed a limited assessment of project impacts that was less complete than a longitudinal study, but indicative nonetheless of project success in certain areas. In the South African assessment, the

⁸ For more on experimental and quasi-experimental methods, see Lucy Creevey and Gary Woller, op.cit.

⁹ For more on selection bias, see Lucy Creevey and Gary Woller, op.cit.

¹⁰ Dean Karlan and Johnathan Zinman, (2006), "Expanding Credit Access: Using Randomized Supply Decisions to Estimate the Impacts," www.microlinks.org/ev_en.php?ID=12529_201&ID2=DO_TOPIC.

researchers conceded that the sampling error meant that the assessment could not be used to prove the impact of relaxing the lending criteria. They argued, nonetheless, that it did provide sufficient evidence to justify a re-examination of the existing lending criteria.

Ensuring Good Performance by the Local Research Partner

Although it is not often emphasized, selecting the local research partner is one of the most important steps in the impact assessment process. Most developing countries have a variety of consulting firms, marketing research firms, research institutes, or universities with experience in local field research. The capabilities of these local researchers, however, can vary considerably. A bad selection can result in higher costs; missed deadlines; greater frustration; poorer quality of work; strained relations with the program, program partners, and donors; questionable results; and, in extreme cases, failure of the research.

Some local researchers are outright incompetent, although unfortunately this may not be discovered until it is too late. In the case of the Azerbaijan project discussed above, for example, costs of the project were raised considerably by the additional expert time required because the local field team did not have the research experience initially claimed. At the same time, competent local researchers may also make mistakes or fail to live up to expectations. Or, the international expert may not adequately communicate what is needed or supervise what the field team does, as was a problem with the baseline research for the AT India project. A related problem occurs when program staff responsible for implementing some part of the research design fails to implement it correctly, whether by mistake, by willful action, or due to a misunderstanding. The latter of these was the case in the South Africa microfinance program cited above.

In any case, problems resulting from mistakes or misdeeds by the local research partner often results in having to redo some part of the work or in discarding the work all together. A recent impact assessment of two USAID tree fruit projects in Kenya suffered from inadequate supervision of data collection during the second round of the longitudinal survey.¹¹ When data from the two survey rounds were matched up, many obvious inconsistencies emerged. This problem could have been avoided if there had been better field supervision, which could have detected many reporting inconsistencies and led to corrective repeat interviews in many cases. This survey also suffered from problems in data processing by the local research partner. Results of these problems included delays, a reduction in the usable sample size, and increased uncertainty about the reliability of the assessment results.

An important part of the solution to these problems is to exercise careful due diligence in vetting and selecting the local research partner.¹² Whereas cost is an important consideration, so is cost-effectiveness. The low cost bidder may not in practice be the low cost option, particularly when the cost of lost time, redoing work, mending strained relationships, or discarding completed work are considered.

Additionally, it is necessary to draft and agree on a carefully delineated and complete Scope of Work (SOW). All the responsibilities and expectations of the local research partner, the program, and the technical advisors should be spelled out in the SOW. It is also important to establish a research manager to serve as the local research partner's

¹¹ Jennefer Sebstad, Don Snodgrass, and Henry Panlibuton (2007), "Impacts of the Kenya BDS and KHDP Projects in the Tree Fruit Value Chain of Kenya," forthcoming.

¹² The approach used by PSD-IAI researchers is to develop a list of local research resources. Sources of information on local research talent include referrals from USAID missions, other in-country development organizations, acquaintances, the internet, etc. Next, conduct a preliminary investigation of potential research partners, such as by visiting their offices and discussing their experience and capabilities. If appropriate, narrow the list down and then invite those on the list to respond to a Request for Proposals. Finally, evaluate the proposals, select the winning proposal, and negotiate a Scope of Work with the winner. The PSD-IAI has provided examples of local research partner RFPs and SOWs at www.microlinks.org/psdimpact.

point of contact and who is charged with managing and monitoring the completion of the field research. This job is frequently assigned to a locally or internationally-based technical consultant.

Finally, good field supervision of data collection is essential if a study is to obtain a reliable data base. In most cases, it is risky to rely entirely on the local research partner for this supervision. It is a good idea for the research manager (and/or a highly reliable assistant) to spend time in the field overseeing implementation, both during the baseline and follow-up research. This might include, for example, participating in the survey pilot test, accompanying some of the enumerators in the first phase of the survey, and working with the field supervisor to see what kind of results are emerging from initial surveys. It may also mean returning when the baseline survey is over and de-briefing the survey enumerators. The degree to which the research manager participates in field work must, of course, take into account budget considerations.

Unanticipated External Events

Even if an impact assessment is well planned, the methodology is sound, and the local research partner is competent, it may encounter outside-project factors that threaten or even wipe out the entire study. One example is the impact assessment undertaken of the craft exporter project in Guatemala. In this case, the baseline research was successfully completed in 2003. The baseline survey included a sample of 1,529 producers of textile, ceramic, wood and leather goods of which 314 were affiliated with the project. The analysis in 2006, however, was based on 56 affiliated producers and 105 non-affiliated textile producers who did not present the same demographic profile as the original textile producers.

At fault here was a major hurricane in 2005 that decimated the central producer population making it impossible to find the original respondents. Many ceased business operations or just left as their businesses and towns were engulfed by the mud slides and disaster resulting from the storm.¹³ The final report carefully delineated project activities and some successful project outcomes reflected in higher sales and incomes of exporters and producers, but it was not possible to prove—or demonstrate without any other likely explanation—the impact of the project activities.

The Azerbaijan project discussed above provides a second case where researchers found it very difficult to overcome outside factors directly affecting the projects success. In this case, there was an outbreak of animal and poultry disease and a serious downturn in the agricultural market. The project, which worked with veterinarians, was directly affected and the results of the impact analysis were distorted.¹⁴ The final report, nonetheless, was able to point out some plausible impacts on veterinarians' business and income attributable to the project.

Conclusion

Most of this paper has been occupied with describing major problems that commonly confront impact assessments of PSD projects. Despite the sometimes discouraging nature of the problems and examples discussed here, the point of this paper is that such problems do occur, and it is better to be upfront about them so that researchers can deal with them directly. In many cases, if the problems are dealt with adequately, the research will still be worth doing and can still produce important findings needed to enable further effective program planning, even if not fulfilling the assessment's original objectives.

A second point that this paper seeks to emphasize is that, when confronted by problems, researchers need to make a clear decision as to whether there is a solution that permits satisfaction of the minimal criteria for a rigorous impact

¹³ Mary Cockram, (2007), "Final Impact Assessment of the ATA-AGEXPRONT Guatemala Craft Sector Export Initiative," forthcoming.

¹⁴ Rees Warne, op cit.

assessment. In some cases an impact assessment will not be possible, but a study of project outcomes and/or a thorough and rigorous analysis of the factors as perceived by participants that influence how and why they have or have not changed their behaviors can be very helpful both for the project or for others considering whether to adopt the project's approach.

One final example is provided by a study of a USAID project in Brazil to increase international market access for three regional sub-sectors: beachwear in Salvador, Bahia; cashew nuts in Barreira, Ceará; and honey in Simplicio Mendes, Piauí.¹⁵ Here the short timeframe of the project (21 months) and the small scale of its activities led the researchers to conclude that a quantitative evaluation in the form of a baseline and follow up survey of participants and a control group would not be appropriate. Instead researchers relied on focus group discussions and in-depth interviews carried out as the project started and then again when it had concluded.¹⁶ The Brazil research team directly confronted the limitations they faced and created a research design that was both rigorous and thorough, although perhaps not ideal.

Whether their approach literally meets the standards of impact assessment research may be beside the point. The results they produce may in fact be more reliable than if another team had chosen to force a quantitative survey approach when circumstances did not permit. What the Brazil team will do at the very least in their final report is provide valuable insights on what the project accomplished and why and whether these findings can plausibly be attributed to the project. Of course, whether settling for such "second best" methods justifies the related costs is another question, and one that all impact researchers and funding organizations must consider.

¹⁵ USAID had also targeted a fourth sector but a tense political situation had put this intervention on hold.

¹⁶Joana Monteiro and Bonnie Brusky, (2006), "Assessing the Impact of Micro and Small-Enterprise Trade-Led Growth Project of USAID Brazil: A Baseline Study," www.microlinks.org/ev_en.php?ID=12718_201&ID2=DO_TOPIC.

U.S. Agency for International Development

1300 Pennsylvania Avenue, NW

Washington, DC 20523

Tel: (202) 712-0000

Fax: (202) 216-3524

www.usaid.gov