IMPACT EVALUATION OF PES PROGRAMS

USAID PES Brief 4

Authors

John Kerr and Rohit Jindal[1]

Need for impact evaluation

PES has many attractive characteristics relative to other conservation approaches provided that transaction costs are low and other favorable conditions apply (see sections 2 and 3 of this Sourcebook). However, ascertaining PES's advantages requires measuring the effect of actual programs in the field. Such impact evaluation can also help in identifying opportunities for further improvements in efficiency of these programs and looking out for other environmental services that can find ready markets. For instance, with the feasibility of selling carbon sequestration services through afforestation and reforestation projects clearly established, researchers are now looking for ways to sell carbon credits from avoided deforestation.

The technical and social complexities of payment for environmental services make impact analysis challenging. Spatial interlinkages, difficulty of perceiving environmental services, the long gestation of benefits, and the multiple objectives of some PES efforts all complicate matters. Many impact studies are therefore either anecdotal or based on a small sample size. Studies that only include PES participants in their sample tend to suffer from selection bias. Further, only some studies have access to baseline information, while many others depend on recall method. This can lead to incorrect inferences about the impact of a PES initiative. The objective of this brief is to suggest some ways of doing impact evaluation studies that can adequately reflect what is going on in the field. This section begins with a quick review of what impact studies should measure.

Impact on environmental services, users, and providers

The overall objective of a PES program is to secure an environmental service by paying for it. Sustainability of a PES initiative is thus directly contingent on establishing the link between the payment and the service delivery. An impact study should therefore be able to measure the level of an environmental service that is available with and without the PES program to establish additionality. For some services such as carbon sequestration, measuring this change is relatively easy. Changes in biomass for a particular tree species are multiplied using known carbon content to calculate the sequestration rate in tons of CO_2 annually. Scenic beauty, on the other hand, is much more difficult to measure, for users vary in their perceptions of it[2]. Biodiversity and watershed services lie between the two. Vegetation type, number of endemic species in an area, and number of different species per unit of area are some of the indicators that can be used to measure changes in biodiversity. Similarly, reduction

[1] Department of Community, Agriculture, Recreation, and Resource Studies, Michigan State University

[2] The difficulty in measuring an environmental service here refers to the challenge in developing an objective scale, rather than the cost of measuring. Measurement costs are covered in detail in Section 3.

in sediment flow, rise in groundwater table, and increase in dry-season flow can be used to verify the impact of a watershed conservation program, depending on what specific service is being sought. In general, the more objective an indicator is, the easier it is to determine the change in the level of the service.

While it is desirable and ultimately necessary to make direct measurements of changes in environmental service indicators (e.g., changes in water flows, water quality parameters, or wildlife numbers) to determine if service providers are obtaining their purchased services, indirect indicators are often used in the short-term for compliance monitoring and measuring implementation progress. Indirect measures are necessary for management purposes because they measure implementation progress during the period in which the ecosystem is being restored and before the ecosystem is capable of delivering the desired ecosystem services. Indirect indicators include measures such as: illegal snares and firearms surrendered, hectares of improved management practices implemented, hectares of riparian zones replanted; forest cover, number of conservation plans agreed to, reductions in pesticide usage, etc.

Many PES programs aim to alleviate poverty by providing payments to poor service providers. In case of the Virilla watershed in Costa Rica, an impact study found that PSA payments led to a 15% increase in the average disposable income of a household. However, a major concern for PES programs is whether poor people can actually participate in a program. For instance, several research studies indicate that, even though Costa Rica's PSA program is beneficial to those poor people who participate, the payments still tend to go disproportionately to the better-educated, wealthier owners of larger farms and forest areas, who are better diversified into non-farm, income-generating activities.

Finally, impact studies should also be carried out to understand the economic value that buyers derive from the environmental service being secured by a PES project. In some cases such as carbon sequestration services, the economic value is easily known by comparing it with international carbon prices. However, in the case of biodiversity conservation or watershed protection, this value needs to be estimated through specific studies. For instance, a downstream dam may gain from reduced silt load due to watershed protection upstream. The economic value can then be calculated in terms of reduced maintenance cost or the increased availability of water for hydroelectricity or irrigation (see USAID PES Brief 3.2, "Valuing Environmental Services," for examples of such studies). Besides such valuation techniques, impact evaluation can also focus on perceptions and attitudes among service users on the level of the environmental service generated through the program. In Ecuador's Pimampiro watershed, for example, many service users felt that it was necessary to protect upstream forests to generate downstream water services, with more than half of the respondents willing to pay more for it.

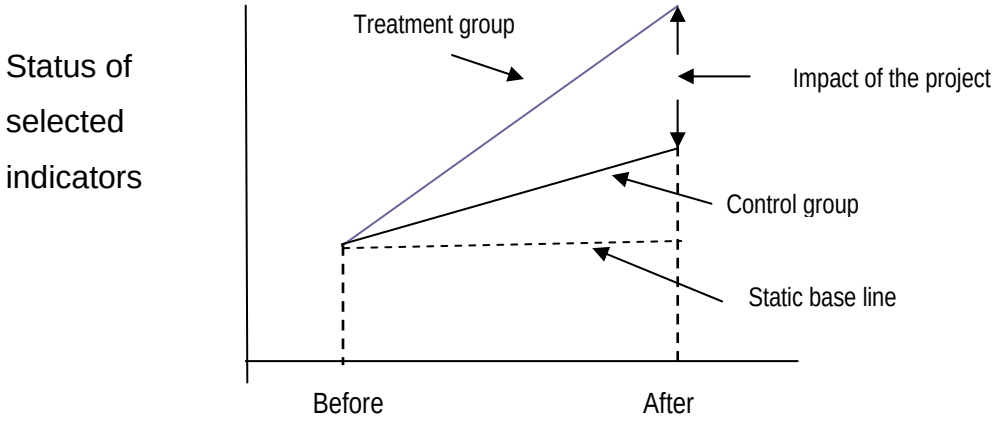
Quantitative evaluation techniques

Quantitative evaluation begins with the premise that the analyst fully understands the nature and determinants of a program's success and can obtain the data needed to measure and relate them statistically. To the extent that it is feasible, quantitative evaluation attempts to attribute changes in various outcome variables to a project intervention or "treatment" and determine whether such effects are statistically significant.

The ideal situation involves an *ex ante* experimental design, complete with randomization of project beneficiaries (e.g., individuals, villages, or project sites) across treatment and control groups. The randomization process has the effect of creating groups that may be considered equal in all attributes, both observed and unobserved, with differences in outcomes attributed to a project. It removes the possibility of sample selection bias, an analytical problem that arises when systematic, preexisting differences between program and non-program locations are correlated with project participation and

the outcome variable of interest.

However, random experiments may not always be possible for PES programs, for choice of sites and participants is often determined by technical criteria. As a result, many evaluations have proceeded with non-randomly determined treatment and control groups, using a variety of quasi-experimental approaches (modeled on experimental approaches). In a before-after study, for example, the evaluator measures the levels of the environmental service before and after an intervention. This requires setting up a base case scenario for indicators that directly relate to the project activities and tracking changes in these indicators to measure the impact of the project (see the figure below).



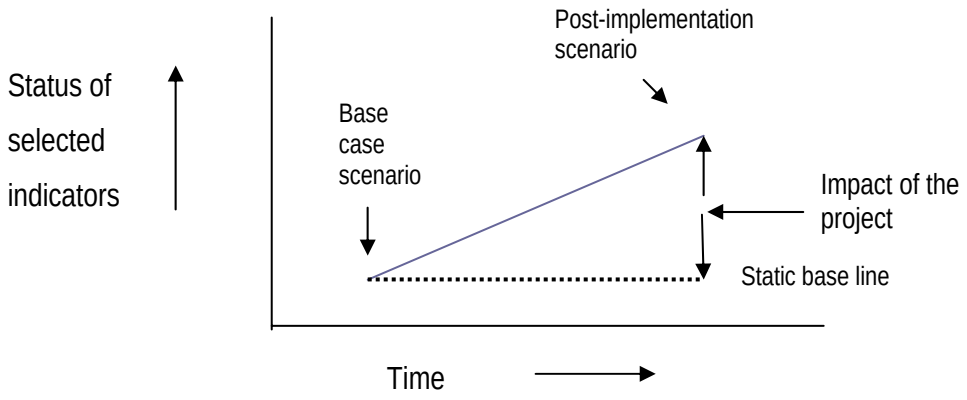
Using a control group to measure project impact

For instance, a study of the Scolel Te carbon sequestration project in Mexico found that discounted benefits for most participants were in the range of -$110 to +$1,700 per hectare. However, this method only calculates the impact with respect to a static baseline, based on the unlikely assumption that there have been no other significant changes during the study period. As a result, it often gives biased results.

Sometimes no baseline data are available, for example, when an evaluation is commissioned after a project has been implemented or the project scope has changed over time. In such cases, researchers can measure only the current state of indicators and must trust respondents to recall the historical status of these indicators. In this case, a with-without design can be useful. To limit sample selection bias, the evaluator must find a control site similar to the treatment sites on as many factors as are hypothesized to affect the outcome.[3] In practice, this is difficult.

Evaluators often suggest a third approach that combines the before-after and with-without approaches. This difference of differences or double difference approach calculates the difference between control and treatment groups at baseline and post-intervention. It has the advantage of "differencing out" any time-invariant unobservable factors that might cause sample selection bias, but it too requires *ex ante* data (see the figure on the next page).

[3] A statistical approach called instrumental variables is used to correct for selection bias in this case. Alternatively, a statistical technique called propensity matching models the probability that each site participates in a project as a function of all observable variables known to affect participation, then matches pairs of participating and non-participating sites that have an equal probability of having been selected for the project. Project impact is estimated as the mean of the differences between all matched pairs on the outcome variable.



Concept of measuring the impact of a project

This approach has been used in the Nhambita Community Carbon Project in Mozambique, where the project has established baselines for the treatment group (Nhmabita) and two control groups (Boa Maria and Munhanganha). The project plans to trace the changes in the three communities over time.

Many studies may lack the time or budget required for careful measurement and must rely on respondents' or investigators' perceptions. For example, one of the impact studies on the Catskill-Delaware watershed protection program to improve water quality in New York measured the perceptions of service providers about their socioeconomic status. The study found that 44.3% of the respondents felt that the program had improved their economic status, while 48.6% felt that it had no effect on them.

Qualitative evaluation approaches

Quantitative approaches provide measured outcomes with statistical tests that support the validity of the findings. But conclusions drawn about a given project are always subject to context-specific conditions. Qualitative methods provide the means by which this context can be understood and may be used to uncover important aspects of a project. Qualitative researchers typically place less emphasis on measurement and more on the process and on understanding the subtle manifestations and determinants of project success, usually by tapping the diverse perspectives of multiple stakeholders. A qualitative analysis is less likely to worry about the applicability of specific outcomes to other project sites, but rather to focus on generalizable 'lessons learned' that may be applied to other projects.

There are many approaches to qualitative evaluation, but they all tend to be flexibly structured and use open-ended questions in an inductive fashion. The objective is not to obtain a numerical estimate of some phenomenon but to develop an in-depth understanding of an issue by probing, clarifying, and listening to stakeholders discuss a topic in their own words. The in-depth nature of the qualitative approach means that a study's scale is usually smaller than in quantitative research, and that the researcher must collect the data rather than hire enumerators. Proponents of a qualitative approach maintain that insights into social processes such as those arising in PES cannot be inferred from measurements of predetermined outcome variables. Rather, the way to understand them is to suspend

one's assumptions about how change occurs and learn from the people who actually experienced a project and its effects. Qualitative evaluators aim to uncover the perspectives of multiple stakeholder groups, learning firsthand about the motivations and dynamics behind decisions and actions taken as a result of a project. More than quantifying outcomes, qualitative evaluations emphasize understanding the processes involved in a project.

For example, in a recent study in India examining the feasibility of linking community forestry projects to international carbon markets, open-ended discussions with community members and NGO officials revealed residents' strong fear that they would lose access to public forest lands if carbon payments were introduced. This demonstrated constraints that a quantitative investigation would have missed.

Mixed methods

Quantitative and qualitative evaluation methods historically have been used separately, but recent years have seen a growing interest in combining the two. The rising interest in combining methods comes from the recognition that both quantitative and qualitative approaches to program evaluation have limitations, and that the strengths of each often compensate the weaknesses of the other. Quantitative approaches are most useful when it is necessary to know the magnitude of a particular effect and when the effect is surely measurable. They are less useful when comparable treatment groups cannot be constructed or when the technical assumptions of the analytical models are not met. Qualitative analysis can provide information about important effects that are not known *a priori*, about the processes that link cause and effect, and about how beneficiaries see the impact.

Mixed methods designs can vary significantly in their structure. Qualitative and quantitative components may be used sequentially, in parallel, or in an integrated fashion. Two main classes of mixed-method designs are 1) a component design and 2) an integrated design. With the component design, qualitative and quantitative methods are used in discrete aspects of a study and are combined only at the level of interpretation or conclusions. Qualitative methods might focus on what actually happened in a project, while quantitative methods might focus on the impact. By contrast, an integrated design mixes methods and allows information collected from one activity to inform data collection for other parts of the study, for example, with ongoing qualitative site visits interspersed into a quantitative evaluation study. Conflicting evidence from the qualitative interviews and the survey would signal that the survey needs improvement. Information from qualitative interviews could be used to revise the survey for later rounds.

This publication was made possible by the United States Agency for International Development and the generous support of the American people for the Sustainable Agriculture and Natural Resource Management Collaborative Research Support Program. The SANREM CRSP, based at Virginia Tech, operates under Cooperative Agreement No. EPP-A-00-04-00013-00.

Global Assessment of Best Practices in Payments for Ecosystem Services Programs

This work was supported at Virginia Tech through USAID Associate Award EPA-A-00-06-00004-00. This work also was supported by USAID through grant LAG-A-00-96-90016-00 to the BASIS CRSP, based at the University of Wisconsin at Madison.

THIS WORK IS INTENDED TO BE A LIVING DOCUMENT THAT WILL BE PERIODICALLY UPDATED AND EDITED. *Updates will be available from the project website. For more information or to send suggestions for changes and additions, see http://www.oired.vt.edu/sanremcrsp/pes or contact Colby at mcolby@usaid.gov*