

CAPRI WORKING PAPER NO. 17

EVALUATING WATERSHED MANAGEMENT PROJECTS

John Kerr and Kimberly Chung



**CGIAR Systemwide Program on
Collective Action and Property Rights**
www.capri.cgiar.org

**Secretariat:
International Food Policy Research Institute
2033 K Street, N.W.
Washington, D.C. 20006 U.S.A.**



AUGUST 2001

CAPRI Working Papers contain preliminary material and research results, and are circulated prior to a full peer review in order to stimulate discussion and critical comment. It is expected that most Working Papers will eventually be published in some other form, and that their content may also be revised.

ABSTRACT

Watershed projects play an increasingly important role in managing soil and water resources throughout the world. Research is needed to ensure that new projects draw upon lessons from their predecessors' experiences. However, the technical and social complexities of watershed projects make evaluation difficult. Quantitative and qualitative evaluation methods, which traditionally have been used separately, both have strengths and weaknesses. Combining them can make evaluation more effective, particularly when constraints to study design exist. This paper presents mixed-methods approaches for evaluating watershed projects. A recent evaluation in India provides illustrations.

KEYWORDS: watershed, natural resource management, project evaluation

TABLE OF CONTENTS

1. Introduction.....	1
2. Some Relevant Characteristics of Watersheds and Watershed Projects.....	3
3. Quantitative and Qualitative Approaches to Project Evaluation	6
4. Case Study: Evaluation of Indian Watershed Projects	18
5. Issues for Future Watershed Evaluations.....	29
References.....	31

ACKNOWLEDGMENTS

This paper is based on a presentation at the Technical Workshop on Watershed Management Institutions, sponsored by the CGIAR System-wide Program on Collective Action and Property Rights, Managua, Nicaragua, March 13-16. Workshop participants, particularly Brent Swallow, Sara Scherr, Ade Freeman and Anna Knox, provided helpful suggestions.

EVALUATING WATERSHED MANAGEMENT PROJECTS

John Kerr¹ and Kimberly Chung²

1. INTRODUCTION

Concern about widespread soil degradation and scarce, poorly managed water resources has led to the spread of watershed management investments throughout Asia, Africa and Latin America (Lal 2000, Hinchcliffe et al. 1999). In India, for example, major rural development programs have been reorganized around a watershed approach, with an annual budget exceeding US\$500 million (Farrington et al. 1999). Despite the growing importance of watershed projects as an approach to rural development and natural resource management, to date there has been relatively little research on their impact.

Clearly, research is needed to ensure that new projects benefit from the positive and negative experiences of their predecessors. Evaluation is difficult, however, due to the social and technical complexity of watershed projects. Typically, watershed project evaluators aim to learn lessons from a limited sample of project sites about how the same projects would perform in other settings. Evaluations usually take either a quantitative or qualitative approach, with the two approaches often viewed as alternatives. International donors such as the World Bank, and research organizations such as the Consultative Group for International Agriculture (CGIAR), tend to favor quantitative evaluations.

¹ Assistant Professor, Department of Resource Development, Michigan State University.

² Assistant Professor, Department of Resource Development, Michigan State University.

Evaluations performed for non-government organizations typically are more qualitative (Hinchcliffe et al. 1999; Farrington et al. 1999).

Evaluation professionals have debated the relative merits of quantitative and qualitative approaches for at least a quarter century (Patton 1997). The 1990s have seen an emerging consensus that both quantitative and qualitative evaluation methods have their own strengths and weaknesses (Patton 1997). Done well, a quantitative approach provides measured outcomes with statistical tests that support the validity of the findings. But even the most positivist evaluators admit that conclusions drawn about a given project are always subject to context-specific conditions (Campbell and Russo 1999). Qualitative methods provide the means by which this context can be understood and may thus be used to expose and examine threats to validity. Campbell and Russo (1999) suggest that social scientists should not limit, trim or change the problems at hand so that they are amenable to scientific precision given the state of the art. Rather, they suggest that social scientists must “stay with (their) problems” and use a larger complement of techniques to improve the validity of the research. This provides a strong rationale for combining approaches to deal with the complexity inherent in projects which must be observed in context (Patton 1997, Henry et al. 1998, Greene and Caracelli 1997), such as a watershed project.

This paper uses an example of an evaluation from India to illustrate the strengths and weaknesses of alternative evaluation approaches and to make the case for using mixed methods. This evaluation was conducted in collaboration between the International Food Policy Research Institute (IFPRI) and the National Centre for Economics and Policy Analysis (NCAP), New Delhi. The study covered dryland

watershed projects operated by government agencies and NGOs in Andhra Pradesh and Maharashtra, two states in India's semi-arid tropical region.

The paper is divided into five sections. Section 2 reviews some distinctive characteristics of watershed development that have implications for impact assessment. Section 3 presents quantitative and qualitative approaches to conducting project evaluation and arguments for combining them. In section 4 the Indian case study is discussed to illustrate the issues, and section 5 concludes with some suggestions about how to promote high quality watershed evaluations in the future.

2. SOME RELEVANT CHARACTERISTICS OF WATERSHEDS AND WATERSHED PROJECTS

A watershed is commonly defined as an area in which all water drains to a common point.³ From a hydrological perspective a watershed is a useful unit of operation and analysis because it facilitates a systems approach to land and water use in interconnected upstream and downstream areas. In dryland areas such as the Indian semi-arid tropics, watershed projects aim to maximize the quantity of water available for crops, livestock and human consumption through on-site soil and moisture conservation, infiltration into aquifers, and safe runoff into surface ponds. In catchment areas of hydroelectric dams, watershed projects typically focus on minimizing soil erosion that deposits sediment into reservoirs and to the maintenance of base flow. In still other

³ This definition corresponds to the definition of "catchment" provided by Swallow, Garrity, and van Noordwijk (1991), and represents the common use of the term in "watershed" projects.

contexts, such as much of North America and Europe, watershed projects focus more on reducing nonpoint source pollution that moves through rivers, streams and drains.

This paper focuses on multiple-use watersheds in hilly or gently sloping areas of developing countries. Such areas are often densely populated and typically contain a variety of land uses, including forests, pastures, rainfed agriculture on sloping lands, and both irrigated and rainfed agriculture in the lowlands. Off-site sedimentation or pollution may or may not be a major issue, depending on the context. It is an important concern in the catchments of river valley projects that provide hydroelectricity and canal irrigation, because sediment can shorten their life span (Hitzhusen 2000). Nutrient transport is also a major concern in river basins that drain into lakes, such as Lake Victoria in East Africa (Swallow et al. 2001). In much of semi-arid India, on the other hand, off-site concerns are typically limited to the local, intra- or intervillage level due to relatively low chemical use and the relative lack of large water bodies.

Watershed projects have numerous distinguishing features that have important implications for both project implementation and impact assessment. These can be divided into at least three categories:

1. *Spatial interlinkages and externalities*: Spatial interlinkages related to the flow of water are inherent in watersheds. Water pollution upstream may harm downstream uses of land and water, while conservation measures upstream may benefit downstream use. Coordination or collective action is often required, which may be difficult because benefits and costs are distributed unevenly. This not only complicates project implementation, but also raises difficulties for evaluation. In particular, since the extent of such complexity will vary by case, a project that works in one location may not work well in another. Subtleties in underlying differences can make it difficult for researchers to understand causal relationships governing project success.

2. *Multiple objectives, dimensions and determinants:* The multitude of project objectives and dimensions and determinants of performance is not surprising given the wide variety of watershed development contexts. Projects may focus on increasing water quantity, improving water quality, reducing sedimentation, or increasing the supply of certain types of biomass, among other things. Some may focus more on organizing people to manage externalities. Project approaches vary with objectives and with local topographic, socioeconomic or cultural conditions. Often they include peripheral activities such as support for agricultural production, marketing, animal husbandry, infrastructure development, or employment generation. Project budgets also vary widely.
3. *Long gestation and difficulty in perceiving project benefits:* Some watershed projects may have short term effects, but all watershed projects have long term impacts, some of which may be difficult to evaluate or even perceive. Soil erosion, for example, is a slow process in many places and the benefits of arresting it may not be recognized easily. Recharging groundwater, stabilizing hillsides through vegetative cover, and increasing soil moisture and organic matter all take time. As a result, it is difficult to know what conditions would have prevailed in the absence of project interventions. Perceiving benefits is particularly difficult where interventions do not raise productivity but merely prevent gradual degradation.

Whether or not a project achieves its objectives depends not only on watershed activities but also a variety of other factors. These may include local agroclimatic conditions, land tenure arrangements, people's willingness and ability to work together to devise arrangements to share benefits and costs, and infrastructure and market conditions that help shape farmers' incentives to manage their land. As a result, it can be difficult to pinpoint the specific contribution of a watershed project in improving land management, and it can be difficult to compare across projects.

Even if impacts are perceptible, it is difficult to assess the economic value of the numerous potential project benefits that do not enter the market. These include such environmental and natural resource improvements as greater abundance and wider

diversity of natural flora and fauna, higher groundwater levels, and lower risk of landslides and flooding, to name a few.

3. QUANTITATIVE AND QUALITATIVE APPROACHES TO PROJECT EVALUATION

Although project evaluation has long been characterized by multiple methodological approaches, until recently evaluators tended to favor either quantitative or qualitative studies (Patton 1997). This is not surprising when one considers the sharply divergent skills required to pursue statistical analysis of project impact, on the one hand, and qualitative assessment of project procedures or changes in beneficiaries' perspectives, on the other. In fact, the difference between the approaches is characterized not just by the methods used, but also by differences in fundamental beliefs about the nature of reality and how claims about this reality are justified. Typically, quantitative studies reflect a positivist view that reality takes a single form that can be perceived and measured objectively. Qualitative approaches, by contrast, reflect a more constructivist view, implying that reality is not separable from individual experiences and that multiple versions of it may exist. From this perspective, an evaluation designed without the flexibility to discover such realities may fail to uncover important aspects of a project (Henry et al. 1998).

The rising interest in combining methods comes from the recognition that purely quantitative and purely qualitative approaches to program evaluation both have limitations, and that the strengths of each often compensate the weaknesses of the other.

The remainder of this section characterizes the two approaches, demonstrates their potential complementarity, and explains the practical basis for combining them.

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. When sample sizes are large enough this methodology is powerful. The randomization process has the effect of creating groups that may be considered equal in all attributes, both observed and unobserved. It removes the possibility of sample selection bias, i.e., an analytical problem that arises when systematic, preexisting differences between program and nonprogram locations are correlated with project participation and the outcome variable of interest (Greene 1999). With no possibility of sample selection bias, the analyst is confident that the outcome is truly a result of the treatment and estimates the program's impact by calculating the difference between the mean of each treatment group and the control. Statistical analysis also requires a sufficient sample size, generated by some form of randomization, rather than a "convenience sample" of a few sites.

An experimental approach is often considered the gold standard of quantitative evaluation. Yet there are reasons why the results of such a study may not extrapolate beyond the projects examined (Manski 1995). First, the conditions of the experimental project site are not likely to be replicated exactly in other sites. Differences in physical, economic and social factors may lead to changes in program outcomes. Second, an experimental program is likely to be carried out differently than the actual program established subsequently. This might occur due to issues of scale. For example, a small experimental program may not affect the market wage or strain the supply of competent program administrators, which would influence the program's effectiveness. Scaling up the program, however, might introduce such constraints and limit performance.

Furthermore, there are many situations in which an experimental approach may not be possible. First, it may be politically or administratively infeasible to randomly assign project sites to treatment groups. Second, many watersheds projects do not deal with sample sizes that make randomization a feasible strategy for study design.

As a result, many evaluations have proceeded with non-randomly determined treatment and control groups. Various approaches have been used, each with their own strengths and limitations. The first is called a "before/after" study. The evaluator measures the levels of outcome indicators in a watershed area before and after an intervention. With this design, the "before" scenario is used as a control against which the effects of the intervention may be compared. This is a fairly weak, but feasible design (Campbell and Russo 1999) that involves the unlikely assumption that there have been no other significant changes during the study period.

This approach often gives biased results as it assumes that without the project, the pre-intervention values of the outcome indicator would have remained the same.⁴ This, however, cannot be known, as it is impossible to observe the same site with and without the intervention. It poses a serious threat to the validity of the findings.

A second approach, a “with/without” design, is useful when no baseline data are available. This is often the case when an evaluation is commissioned after a project has been implemented. As such, randomization is impossible and sample selection bias is likely. To reduce this threat, the evaluator must find a control site that is similar to the treatment sites on as many factors as are hypothesized to affect the outcome. However, in practice, sites are likely to vary in almost an infinite set of ways, and evaluators try to match sites on only those factors that suggest likely threats to validity.

Clearly, decreasing sample selection bias depends on the extent to which the evaluator is able to create comparable treatment and control groups. Jalan and Ravallion (1998) used a statistical technique called propensity matching to match on the basis of multiple factors. This involves modeling the probability that each site participates in a project as a function of all observable variables known to affect participation, and then matching 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.

Such approaches to with/without analysis may succeed in creating treatment and control groups that are equivalent in terms of observable characteristics, but they cannot

⁴ For example, this approach will not measure any benefit from a project that arrests degradation of the resource which would otherwise have taken place without the project.

control the effects of *unobservable* characteristics. To the extent that some factors that determine program placement are unknown, selection bias may persist (Baker 2000). Given this problem, it is not surprising that evaluators often suggest a combination of 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 (Baker 2000). But it also requires the assumption that these unobservable factors have not changed during the study period. In addition, the evaluation must be commissioned *ex ante* as data on participants and non-participants are required before and after the intervention.

All of the above approaches have been modeled after the scientific tradition of experimental design and are thus termed “quasi-experimental.” Social scientists have developed another approach to deal with the inherent problems of sample selection bias when quasi-experimental designs are infeasible or insufficient. Rather than comparing treatment and control groups, a statistical technique known as instrumental variables is used to remove the bias introduced by sample selection bias (Greene 1999). Typically, a two-stage model is used; one equation models the probability that a given observation is selected (or self-selects) for a given program. A second estimates the outcome in question, replacing the endogenous treatment variable with its predicted value. This process adjusts for the selection bias if, 1) exogenous “instruments” can be found that are significant determinants of project participation but do not directly affect the outcome of interest conditional on participation and 2) the participation model is valid.

The instrumental variable procedure carries the advantage that impact evaluations may be conducted *ex post*, as long as appropriate data exist for the non-participating sites. Its disadvantages are 1) the estimated effect is highly dependent on the validity of the chosen instruments and 2) appropriate instruments are often difficult to find. In cases where inappropriate instruments are used, the bias introduced by the two-step procedure can be worse than the bias it was attempting to correct (Bound et al. 1995).

Aside from issues of design, the specification of outcome variables presents yet another problem for quantitative watershed evaluations. As mentioned above, measuring improvements in natural resource conditions is difficult. Many studies lack the time or budget required for careful measurement and must rely on respondents' or investigators' perceptions. Even where measurement is possible, the data it provides may be of limited use. For example, recent research shows that traditional runoff plots are unreliable for extrapolating differences in soil erosion across management practices within a site, because these differences may be dwarfed by those across sites that vary in exposure, slope or soil conditions (Schreier 2000). The long gestation and uneven, uncertain spatial distribution of project impact compound the measurement difficulties.

Cost-benefit analysis

Cost-benefit analysis has long been the method of choice in economic appraisal of agricultural development and irrigation projects. Cost-benefit analysis focuses on assessing whether a project yields net societal benefits (Gittinger 1982). Cost-effectiveness analysis is similar but it estimates only the costs of alternate approaches of achieving a given objective. Cost-benefit analysis aims to evaluate costs and benefits that occur with a project and compare them to what would happen without the project.

Obviously the without-project outcome cannot be observed and must be estimated. This involves estimating adoption rates and trying to determine to what extent they can be attributed to the project, and then estimating the effect of adoption on technical relationships, prices and incomes.

This approach is complex enough when the task at hand is to measure the costs and benefits of a project that develops a new technology, such as a new variety of grain, or that introduces irrigation to a dryland area. In these cases the adopters are easily distinguished from nonadopters and adoption can be attributed to the project. In addition, measuring changes in production, while never perfect, is reasonably straightforward.

In a natural resource management project, however, the task is much more complicated (Traxler and Byerlee 1992). First, a natural resource management objective may be achieved by many different means and evaluators must not mistakenly attribute to a project gains that accrue from independent actions. In India, for example, some projects introduced contour vetiver grass hedges to conserve soil and moisture, but this approach is not necessarily more effective than traditional grass strips on the lower boundaries of small plots (Kerr and Sanghi 1992; RAU 1999). Many farmers used the traditional practices without help from a watershed project, and evaluators who were not aware of these practices exaggerated project impact.

Second, many projects promote existing practices (such as grass strips or stone or earthen barriers), and it is difficult to estimate how many more farmers use them because of the project.

Third, as with other quantitative evaluation methods, cost-benefit analysis depends heavily upon the accuracy of the data used and this raises the problems introduced above.

Fourth, the difficulty of assigning prices to environmental services poses obvious challenges to cost benefit analysis. Environmental economists have developed ways to estimate the value of such unpriced services, but data limitations and uncertainties may limit their applicability to the case of developing country watershed projects. (Cost-effectiveness analysis avoids the need to attach values to environmental benefits.)

Finally, even if all costs and benefits could be identified and valued, cost-benefit and cost-effectiveness analysis would give only a single assessment of overall project performance. Watersheds, however, consist of multiple users who are affected differently by the project. A favorable benefit:cost ratio could mask uneven distribution of benefits, yet those who do not benefit may be in a position to undermine the project. In this case a project with high aggregate net benefits may not be sustained, making projected benefits illusory.

To summarize, there are clearly multiple challenges associated with using quantitative evaluation methods for evaluation of watershed projects. Most challenges are introduced by the fact that watershed projects are not amenable to the same controlled conditions that bestow power and simplicity on the analysis of data collected in the experimental sciences. Specifically, the advantage of clearly interpretable outcomes is tempered by threats to validity resulting from unreliable data and models that require strong assumptions. If the data or model assumptions are inaccurate, statistical findings may not be internally valid (correct within the sample), let alone externally valid. Of

course, it may be possible to obtain more accurate data, but only at the cost of more time or money, neither of which may be available. Specialized econometric techniques may compensate for some weaknesses in study design, but they too require strict assumptions. Also, they are beyond the understanding of many end-users and some argue that the lack of transparency will lower an evaluation's credibility among them (Patton 1997).

The important point is that no approach is perfect. The evaluator must address the threats to validity implied by the assumptions associated with each. This in turn depends on the evaluator's skills, the project's attributes, the resources and data available to the study, and the timing of the evaluation relative to project implementation.

QUALITATIVE EVALUATION APPROACHES

In contrast to quantitative analysts, qualitative researchers typically place less emphasis on measurement and more on context and on understanding the subtle manifestations and determinants of project success, usually by tapping the diverse perspectives of multiple stakeholders (Cronbach 1982, Henry et al. 1998). A qualitative analysis is less likely to worry about the generalizability of specific outcomes to other project sites, but rather to focus on generalizable 'lessons learned' that may be applied to any kind of project.

There are many diverse approaches to qualitative evaluation (Patton 1990). In general, however, a qualitative approach tends to be flexibly structured and uses open-ended questions in an inductive fashion. The data collection process allows for the emergence of important dimensions not previously known to the researcher. 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 talk about a topic in their own words. The process is iterative in that the researcher keeps trying to clarify his/her understanding of a phenomenon. He/she may therefore ask unscripted follow-up questions to probe for a clearer, more nuanced understanding. Or he/she may return later to clarify a point that came up in the interview or to validate information collected in an interview with another individual.

Qualitative researchers are comfortable asking respondents to give their own interpretation of “why” and “how” something happens. They are more interested in fully understanding why individuals behave the way they do in a given situation, given its unique circumstances, rather than generalizing an outcome across numerous cases. They use theory to provide a conceptual framework for starting their work, but they constantly update their understanding of the situation as more information is collected. This process generates an explanation that is grounded in the context studied.

The in-depth nature of the qualitative approach means that a study’s scale is usually smaller than that found in quantitative research. Proponents of a qualitative approach maintain that insights into social processes such as those arising in watershed management cannot be inferred from measurements of pre-determined outcome variables. Rather, the way to understand them is to suspend one’s assumptions about how change occurs and instead learn from the people who actually experienced a project and its effects. As such, qualitative evaluators aim to uncover the perspectives of multiple stakeholder groups, learning first hand about the incentives, motivations, and dynamics behind decisions and actions taken as a result of a project. Qualitative

evaluations, therefore, emphasize understanding the processes involved in a project more than quantifying outcomes.

As with quantitative analysis, sampling issues in qualitative research also raise questions about biases in data. While quantitative researchers use random sampling whenever possible (and statistical fixes when it is not), qualitative researchers use several strategies to increase the internal validity of their findings. Of these, triangulation, the method of using different subjects, settings, or data collection methods to gain a better assessment of the soundness of a given finding, is the most well known. Qualitative researchers also use member checking, a method of systematically soliciting feedback from respondents on the data collected and tentative conclusions. Maxwell (1996) cites this as the single most important method available to ensure that the researcher has not misinterpreted what has been said or observed. Qualitative researchers also search for discrepant or negative cases to falsify a proposed conclusion. Finally, like quantitative researchers, they rely on their judgment, their caution, and their emerging understanding of the context to estimate the seriousness of any given threat to validity.

A final difference in research approach concerns the researcher's role in data collection. Typically, quantitative researchers analyze data that someone else has collected, at most visiting the study area to gain some understanding of the context. In qualitative research, on the other hand, the processes of data collection and data analysis are intertwined, with the researcher's interpretation of data that is collected one day affecting decisions about data collected the next. Thus, qualitative data collection and analysis become inseparable; as such researchers collect much of the data themselves, rather than relegating this task to field assistants.

MIXED METHODS EVALUATION DESIGNS

It is clear that different approaches to evaluation carry different requirements, assumptions, strengths and weaknesses. There is a growing acceptance that very different approaches to evaluation can contribute complementary insights. Quantitative approaches may be particularly 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 impacts that are not known *a priori*, about the processes that link cause and effect, and about how beneficiaries see the impacts.

Researchers use mixed methods designs for various reasons. Patton (1997) represents the pragmatic methodologists -- those who suggest mixing methods opportunistically, using whatever approach is best suited for a given task. As an example, Datta (1997) cites a case in which the United States Agency for International Development (USAID) planned to evaluate a child survival project in Indonesia. Due to data, time, and staff limitations, the evaluators chose to do a mixed-methods evaluation using secondary data sets, existing documents, and qualitative interviews. With less than three weeks on-site, the team designed a study that combined data from various sources and optimized various trade-offs given the constraints. The authors took particular care to use the complementary types of data to rule out plausible rival hypotheses.

Mixed methods designs can vary significantly in their structure. Qualitative and quantitative components may be used sequentially, in parallel, or in an integrated fashion (Tashakkori and Teddlie 1997). Caracelli and Greene (1997) suggest two main classes of

mixed-method designs: 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. Such studies tend to have a more pragmatic orientation since the design presents little opportunity for tacking between different paradigms. In the example presented by Datta (1997), a quasi-experimental study was used to answer one evaluation question (“What were the impacts on infant and child mortality?”), while qualitative document analysis and interviews were used to answer another (“How were the activities implemented?”).

By contrast, an integrated design mixes methods and allows information collected from one activity to inform data collection for other parts of the study. Mark et al. (1997) describe a study in which on-going qualitative site visits were interspersed into a quantitative evaluation study. The authors obtained conflicting evidence from the qualitative interviews and the survey and used this discrepancy as a signal that the survey had a problem. Using the information provided by the qualitative interviews, they revised the survey for later rounds. In short, conflicting evidence suggested areas that were not yet well understood. They also claim “productive dialectics sometimes occur and sometimes do not.” They suggest designing a mixed-methods evaluation in a way that 1) allows such a dialectic to emerge and 2) that employs the relative strengths of the different methods.

4. CASE STUDY: EVALUATION OF INDIAN WATERSHED PROJECTS

The IFPRI-NCAP watershed evaluation study in India illustrates many of the issues introduced in the previous sections. The study, conducted in 1996-98, was part of

a larger effort coordinated by the World Bank (WB) and the Indian Council of Agricultural Research (ICAR) -- the research arm of the Ministry of Agriculture (MoA) - - to identify priorities for investing in predominantly rainfed agricultural areas. The study focused on Maharashtra, the state with the most experience in watershed development, and Andhra Pradesh, a state likely to be targeted for a rainfed agricultural development loan.

Despite the large budgets devoted to watershed development, reliable evaluation studies were scarce at the time the study was initiated. Some early studies indicated high adoption rates of soil and water conservation practices and favorable benefit-cost ratios (IJAEE 1991). However, these studies focused on heavily supervised projects with subsidies of 90-100% awarded to adopters of the prescribed packages. As such, the estimates of adoption rates were not meaningful. Also, the benefit-cost studies were conducted before the actual outcomes could be known. They estimated net project benefits using yield impacts based on experimental data and assuming adoption and maintenance rates by farmers (e.g. Singh et al. 1989). *Ex post*, however, some evidence suggested that many farmers abandoned watershed measures once the project subsidies ended (Kerr and Sanghi 1992). Taken together, these factors suggested that many of the early, favorable evaluations were overly optimistic.

On the other hand, there was detailed documentation of a small number of highly successful projects that highlighted innovative social organization arrangements or the influence of exceptional leadership in addition to technical interventions (e.g. Chopra et al. 1990). Many NGOs gave reports of their own successful watershed development initiatives, and while there were undoubtedly many favorable projects, it is also likely

that these reports focused mainly on the best cases and gave less attention to the problems they faced.

A MIXED METHODS APPROACH

IFPRI, NCAP and the WB were primarily interested in economic analysis that would compare multiple projects and identify which of the many approaches to watershed development in India were most successful. It would also capture the role of exogenous factors, such as infrastructure, in determining the outcomes of interest: agricultural productivity, natural resource management and poverty alleviation.

The terms of reference explicitly called for a combination of quantitative and qualitative analysis, but the composition of the research team predisposed it to make the quantitative component its primary concern. The principal investigators from IFPRI and the WB managers and advisors for the study were all economists. All of them were knowledgeable about Indian agriculture, but none were professional evaluators or had extensive experience with qualitative methods. The ICAR officials overseeing the project included agricultural scientists who also were predisposed towards a quantitative study modeled on the scientific approach.

Originally, researchers intended to use a sequential mixed-methods approach. In practice, however, the project time frame did not allow the qualitative data to be collected and analyzed before the quantitative study was implemented. ICAR and the WB were under pressure to complete the studies within eighteen months since a large loan for rainfed agriculture was contingent on their findings. The logistical difficulties of developing a sampling frame for the quantitative study reduced the time available to

analyze and interpret the qualitative data *ex ante*. As such, the mixed-methods design was effectively a parallel, components design.

STUDY DESIGN

The village was selected as the unit of analysis since most Indian watershed projects operate at the village level and the people affected by the projects are organized in villages. The quantitative component was conducted as a “with and without” design, covering five project categories. These included four different treatment groups -- two types of government projects, NGO projects, government-NGO collaborative projects -- and a control group of nonproject villages (see Table 1).

Table 1--Project Categories in the Evaluation of Indian Watershed Projects

1. ***Ministry of Agriculture (MoA)***: projects that focus primarily on technical aspects of developing rainfed agriculture.
2. ***Ministry of Rural Development (MoRD)****: Engineering-oriented projects that focus on water harvesting through construction of percolation tanks, contour bunds, and other structures.
3. ***Non-government organizations (NGOs)***: projects that typically place greater emphasis on social organization and less on technology relative to the government programs.
4. ***NGO-Government collaboration***: projects between government and non-government organizations that seek to combine the technical approach of government projects with the NGOs’ orientation toward social organization.
5. ***Control***: villages with no project.

All of these project categories are discussed in detail in Kerr (2000).

* This study did not include villages under the new guidelines of the Ministry of Rural Development, which called for more attention to social organization. The projects were just getting underway at the time of the data collection for this study, so it was too soon to include them.

To avoid choosing only conveniently located sites or success stories, researchers generated a stratified random sample from a census of villages where watershed projects were concentrated. Ultimately 86 villages, stratified by the five project categories, were sampled from a frame of over a thousand villages in the two states. While it was important to randomly sample the sites to be studied, generating the census of watershed projects was particularly time-consuming because such information was not available from official records. The quantitative analysis covered all the sampled villages, while the qualitative analysis focused on a randomly selected subsample of 29 of those villages.⁵

This study encountered many of the challenges cited in Section 3. As such, its design reflects the constraints imposed upon the research team. To start, there was no baseline data on the performance criteria that were of interest to the evaluation team. As such, multiple indicators were used to assess project performance, some of which were based on respondents' perceptions. Respondents' recall was used for indicators that could be defined in terms of an easily observed, discrete change between one period and the next, such as adoption of new varieties, changes in infrastructure, and ownership of assets. Table 2 shows how performance criteria of interest were operationalized into indicators.

⁵ Watersheds fall within village boundaries in all project categories except the Ministry of Agriculture, in which a watershed covers multiple villages.

Table 2—Ideal and Operationalized Indicators of Performance

Performance criteria	Ideal indicators¹	Operational indicators used in this study
soil erosion	- measurement of erosion and associated yield loss	- visual assessment of rill and gully erosion (current only)
measures taken to arrest erosion	- inventory, adoption and effectiveness of SWC practices	- visual assessment of SWC investments and apparent effectiveness (current only) - adoption of conservation-oriented agronomic practices - expenditure on SWC investments
groundwater recharge	- measurement of groundwater levels, controlling for aquifer characteristics, climate variation and pumping volume	- approximate change in number of wells - approximate number of wells recharged or defunct - change in irrigated area - change in number of seasons irrigated for a sample of plots - change in village-level drinking water adequacy
soil moisture retention	- times series, intrayear and interyear variations in soil moisture, controlling for climate variation	- change in cropping patterns - change in cropping intensity on rainfed plots - relative change in yields (higher, same or lower)
agricultural profits	- net returns at the plot level	- net returns at the plot level, current year only
productivity of nonarable lands	- change in production from revenue and forest lands (actual quantities) - wildlife habitat	- relative change in production from revenue and forest lands (more, same or less than pre-project) - extent of erosion and SWC on nonarable lands - change in wildlife and migratory bird populations
household welfare	- change in household income and wealth - nutritional status	- perceived effects of the project on the household - perceived change in living standard (better, same, worse) - change in housing quality - change in percentage of families migrating - perceived changes in real wage and availability of casual employment opportunities (higher, same, lower)

¹All ideal indicators would be collected both before and after the project.

Second, a lack of secondary data on the sites from the initial census precluded the use of propensity matching to construct control and treatment groups. Rather, the groups were stratified by project type and topography of the project site (hilly vs. flat).

Third, the project sites were not originally assigned through a random process, so sample selection bias was an issue. Site-selection criteria differed by project type. MoA programs, for example, favored more accessible locations to facilitate demonstration visits by officials and people from other villages (Government of India 1992). These villages had better access to markets, perhaps raising the incentive to invest in rainfed agriculture. NGOs, on the other hand, favored remote villages with less access to markets and government services. Some NGOs also selected villages where people had already demonstrated the ability to work collectively. An instrumental variables approach was employed to account for the problem of sample selection bias.

The qualitative component aimed to augment the quantitative investigation in two ways. First, it focused on learning people's key concerns and how projects affected them. Second, it sought to identify alternative indicators of some of the performance measures collected in the quantitative data. The approach involved group interviews and focus group discussions with specific interest groups in the village, such as farmers with irrigated land, farmers without irrigation, landless people (often herders), and people with low castes. Men and women were interviewed separately. This approach helped gain information about the distribution of project benefits and costs. The sampling of groups within the village was opportunistic, and the discussions followed a common framework in every village.

Given the limitations of the study, the evaluation team recognized that it would be important not to depend on any single statistical estimate in drawing conclusions (Manski 1995). Rather, it would be important to consider various threats to validity posed in the quantitative analysis and to triangulate these findings against the data collected through the qualitative components. This study, therefore, represents a pragmatic, mixed-method evaluation.

FINDINGS

Only an overview of the findings is presented here; detailed results are available in Kerr (2000). Both the quantitative and qualitative analyses gave support to better performance by those projects with an NGO component. This was true for a range of performance categories such as soil conservation on drainage lines and common pasture lands, adoption of new crop varieties, and net returns to cultivation. Performance in government project villages, on the other hand, often was not significantly different from that in control villages.

According to the analysis, NGO and NGO-government collaborative projects appear to have been more successful in promoting collective action, which was manifest in arrangements to protect common pasture lands and drainage lines. This may be in part because they selected villages predisposed to collective action, but the same result was obtained even when econometric techniques were used to control for sample selection bias. The fact that NGOs devoted at least a year, and often several years, to social organization while government projects rarely devoted more than a month, makes this finding unsurprising. Details from qualitative interviews about how some of the NGOs promoted social organization, and the kinds of institutional arrangements they helped

establish, also support this finding. In Andhra Pradesh, for example, some NGOs worked for years to help specific interest groups in a village organize themselves, creating a capacity for self-determination among even the poorest and politically weakest community groups. They facilitated negotiations among different groups and helped enforce agreements. Such attention to social organization was unheard of in the government programs included in the study.

NGO and NGO-government collaborative projects also performed as well as or better than government projects in promoting adoption of improved agricultural technology and generating higher agricultural income. This result was unexpected, because the NGO projects focused less on agriculture, and they operated in villages with apparently less favorable conditions for agricultural intensification. One possibility is that because they began from a lower technological base, their more rapid adoption of improved technology may be simply a process of catching up. Another reasonable explanation is that many of the NGOs helped promote agricultural production indirectly, for example by putting pressure on government extension services to focus on a particular village or lobbying for infrastructure improvements. In some places they obtained market information from distant cities and then helped farmers arrange transport to sell their produce in locations with higher prices. Information about such approaches came only through qualitative interviews.

The qualitative data were particularly helpful for understanding the extent to which different groups of people were involved in establishing project priorities and their perceptions of projects' distributive impacts. For example, qualitative interviews with landless people in many of the Maharashtra villages revealed that they had little say in

project decisions and felt harmed by the projects. This was true for both government and NGO projects that aimed to close common lands to grazing, a livelihood on which many landless people were dependent. The landless could be excluded from this decision because most of the Maharashtra projects required that villages vote to determine whether to accept a project. A 70% majority was needed to initiate these projects, and in most villages the landless population was too small to mount a successful opposition. Such findings illustrate the importance of understanding local institutions and the power that institutional processes may have in determining the distribution of project outcomes.

For some indicators, the quantitative analysis did not detect impact by any projects. Expanding irrigated area is an example: changes in irrigated area showed no association with project category or the extent of project investment. The most likely reason is poor and missing data. Probably the most important factors in determining changes in irrigated area are the characteristics of the aquifer and the amount of rainfall, but no appropriate information was available. Also, changes in irrigation due to watershed development may have been minor; for example, water levels might have been slightly higher in wells under watershed projects, but the difference may have been too small to affect irrigated area or cropping patterns. Qualitative investigation suggested that farmers perceived water harvesting structures in drainage lines to be effective in raising groundwater levels, but also that they often could not distinguish between the effects of water harvesting efforts and changes in rainfall.

The study's final report was delivered to ICAR and World Bank officials in 1998 and presented in government-sponsored workshops. Its focus on quantitative data helped make it useful for Indian policymakers. The finding of poor performance of government

projects was unpopular, but the quantitative results gave it credibility that purely qualitative results would not have enjoyed. The fact that the qualitative findings reinforced the quantitative results was important given the imprecision of the quantitative analysis: in isolation, both the quantitative and qualitative results would have been less credible.

Timeliness of the results was also important. Given the constraints placed on the study, the research team concluded that there would be little benefit to engaging in a more statistically complex study design. Of particular note, the study was commissioned *ex post* and policymakers were anxious to apply the results to their decisions about future WB loans. As such, investing twice as much time collecting more complicated forms of data or conducting higher levels of econometric tests was unimportant to the end-uses. Instead, the report contained fairly simple statistical corrections for sample selection bias and concentrated on providing a best-case evaluation given the constraints.

We believe that this choice made sense for the situation. Within a year of submission of the final report, the MoA decided to reorganize its watershed projects on a much more participatory approach that includes a greater role for NGOs. It would be unrealistic to attribute this change in policy exclusively to the IFPRI-NCAP evaluation, because the Ministry of Rural Development (MoRD) had already initiated such a change a few years earlier, and many other voices pointed to the need for greater orientation to social organization in MoA programs. Still, it is likely that the evaluation did play a role. As one of very few quantitative studies of project performance, it reinforced the other voices that favored more participatory approaches oriented toward social organization. Islam and Garrett (1997) argue that policy analysis studies are likely to

have the greatest impact when they are conducted at a time when they lend support to ideas that have already gained some acceptance, when policymakers are open to the idea of policy change, and when the policymakers are kept informed of the progress of the evaluation.

5. Issues for Future Watershed Evaluations

As the CGIAR and other international development organizations become more involved in evaluating watershed projects (and other research and development activities), they have much to gain by embracing mixed methods approaches. To date the CGIAR institutes have favored quantitative analysis, and the quality of their work is high. There is no reason for them to abandon this work; rather, the idea is to further strengthen it by adding a qualitative research component to yield complementary information.

The IFPRI-NCAP watershed evaluation study demonstrates the advantages of employing mixed methods as well as some of the practical constraints to achieving an ideal study. It has lessons for future mixed-methods evaluations that function in the real world, where data are inadequate and decision makers cannot wait years for results. Operating with a lack of baseline data and lack of access to precise indicators of performance, the investigators performed a best-case quantitative analysis and augmented it with insights generated from qualitative work. However, the qualitative investigation was less thorough than desired, because logistical challenges related to the quantitative data collection limited the time that principal investigators could spend in the field focusing on the qualitative components. This is a common problem with mixed-methods studies in which one approach takes precedence over the other. It represents a lost

opportunity in terms of the synergies that might have been generated had findings from both the quantitative and qualitative approaches been available to inform each other. This experience helps demonstrate the tradeoff between the depth and scope of a mixed-methods study: sharpening the focus of the quantitative component may have enabled the principal investigators to spend more time engaged in the qualitative investigation. Were the study to be conducted again under identical circumstances, this would be the best way to proceed.

A second lesson is that future evaluations may benefit from focusing not simply on final outcomes but also on the processes that lead to those outcomes. This is particularly important in watershed development, where specific technical interventions will vary by site but the processes of technology assessment and social organization might be similar.

Third, including the expected users of evaluations in the design process is another good practice and a good reason to incorporate qualitative methods that may be relatively easy to understand or that may provide specific examples to support important points. The International Institute for Environmental and Development (IIED), for example, engaged watershed development agencies in self-evaluation studies so that they would think critically about their own work (Hinchcliffe et al. 1999). They claim it is likely that many of them put their evaluation findings to work in their projects. Finally, participatory evaluations that include project participants, not just the implementing agencies, have the potential to generate greater understanding of project impacts and to provide local people with greater influence over how projects operate (Cousins and Whitmore 1998).

REFERENCES

- Baker, Judy. 2000. *Evaluating the impact of development projects on poverty: A handbook for practitioners*. Washington: World Bank.
- Bound, John, Jaeger, David A., and Baker, Regina M. 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous variable is weak. *Journal of the American Statistical Association*, 90(430) 443-450.
- Campbell, Donald T. and M. Jean Russo. 1999. *Social experimentation*. Thousand Oaks: Sage Publications.
- Caracelli, Valerie and Jennifer Greene. 1997. Crafting mixed-method evaluation designs. In *Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms*, ed. Jennifer Greene and Valerie Caracelli. New Directions for Evaluation, No. 74. San Francisco: Jossey-Bass.
- Chopra, Kanchen, Gopal Kadekodi, and M.N. Murty. 1990. *Participatory development, people and common property resources*. New Delhi: Sage Publications.
- Cousins, J. Bradley, and Elizabeth Whitmore. 1998. Framing participatory evaluation. In *Understanding and practicing participatory evaluation*. ed. Elizabeth Whitmore, New Directions for Evaluation, No. 80. Jossey-Bass: San Francisco.
- Cronbach, Lee. 1982. *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass.
- Datta, Lois-Ellin. 1997. A pragmatic basis for mixed-method designs. In *Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms*, ed. Jennifer Greene and Valerie Caracelli. New Directions for Evaluation, No. 74. San Francisco: Jossey-Bass.
- Farrington, John, Cathryn Turton and A.J. James, eds. 1999. *Participatory watershed development: Challenges for the twenty-first century*. New Delhi: Oxford University Press.
- Gittinger, J. Price. 1982. *Economic analysis of agricultural projects*. Baltimore: Johns Hopkins.
- Government of India. 1992. WARASA: National Watershed Development Project for Rainfed Areas (NWDPR) Guidelines (3rd edition). New Delhi: MoA.

- Greene, Jennifer, and Valerie Caracelli. 1997. *Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms*. New Directions for Evaluation, No. 74. San Francisco: Jossey-Bass.
- Greene, William. 1999. *Econometric analysis*. Englewood Cliffs, NJ: Prentice Hall.
- Henry, Gary, George Jules, and Melvin Mark. 1998. Realist evaluation: An emerging theory in support of practice. *New Directions for Evaluation*, No. 78. San Francisco: Jossey-Bass.
- Hinchcliffe, Fiona, John Thompson, Jules Pretty, Irene Guijt and Parmesh Shah, eds. 1999. *Fertile ground: The impacts of participatory watershed management*. London: Intermediate Technology Publications.
- Hitzhusen, Fred. 2000. Economic analysis of sedimentation impacts for watershed management. In *Integrated watershed management in the global ecosystem*, ed. Lal, Rattan. Boca Raton, Florida: CRC Press.
- Indian Journal of Agricultural Economics. 1991. Subject 1. *Watershed development*. Vol. XLVI (July-September), pp 241-327.
- Islam, Yassir and James L. Garrett. 1997. *IFPRI and the abolition of the wheat flour ration shops in Pakistan: A case study on policymaking and the use and impact of research*. Impact assessment discussion paper No. 1. Outreach Division, Washington, DC: International Food Policy Research Institute.
- Jalan, J., and Martin Ravallion. 1998. Transfer benefits from welfare: a matching estimate for Argentina. Mimeo. Washington, DC: World Bank.
- Kerr, John. 2000. *The role of watershed projects in developing India's rainfed agriculture*. EPTD Discussion Paper No. 68. Washington, DC: International Food Policy Research Institute.
- Kerr, John, and NK Sanghi. 1992. *Indigenous soil and water conservation in India's semi-arid tropics*. Gatekeeper Series no. 34, Sustainable Agriculture Program, London: International Institute of Environment and Development.
- Lal, Rattan, Ed. 2000. *Integrated watershed management in the global ecosystem*. Boca Raton, Florida: CRC Press.
- Manski, Charles. 1995. *Identification problems in the social sciences*. Cambridge: Harvard University Press.
- Mark, Melvin M., Feller, Irwin, and Button Scott. 1997. Integrating qualitative methods in a predominantly quantitative evaluation: A case study and some reflections. In: Greene Jennifer and Caracelli Valerie (eds). San Francisco: Jossey-Bass, Inc.

- Maxwell, Joseph. 1996. *Qualitative Research Design: An interactive approach*. Applied Social Research Methods Series, Volume 41. Thousand Oaks: Sage Publications.
- Patton, Michael Quinn. 1997. Utilization-focused evaluation. 3rd Edition. Thousand Oaks: Sage Publications.
- RAU. 1999. Impact Evaluation Report: World Bank Integrated Watershed Development Project (Plains), Rajasthan. College of Technology and Agricultural Engineering, Rajasthan Agricultural University.
- Schreier, Hans. 2000. Personal communication based on research in Nepal.
- Singh, R.P., K.P.R Vittal, S.K. Das, N.K. Sanghi. 1989. Watershed technology stabilizes yields in Andhra Pradesh. *Indian Farming* 39 (9).
- Swallow, Brent, Dennis Garrity, Meine van Noordwijk. 2001. *The effects of scales, flows and filters on property rights and collective action in catchment management*. CAPRI Working Paper 16. <http://www.cgiar.org/capri/pdf/capriwp16.pdf> Washington DC: International Food Policy Research Institute.
- Tashakkori, Abbas, and Charles Teddlie. 1997. *Mixed methodology: Combining qualitative and quantitative approaches*. Applied Social Research Methods Series, Volume 46. Thousand Oaks: Sage Publications.
- Traxler, Greg, and Derek Byerlee. 1992. Economic Returns to Crop Management Research in a Post-Green Revolution Setting. *American Journal of Agricultural Economics*, August 1992, pp 573-582.

CAPRI WORKING PAPERS

LIST OF CAPRI WORKING PAPERS

- 01 *Property Rights, Collective Action and Technologies for Natural Resource Management: A Conceptual Framework*, by Anna Knox, Ruth Meinzen-Dick, and Peter Hazell, October 1998.
 - 02 *Assessing the Relationships Between Property Rights and Technology Adoption in Smallholder Agriculture: A Review of Issues and Empirical Methods*, by Frank Place and Brent Swallow, April 2000.
 - 03 *Impact of Land Tenure and Socioeconomic Factors on Mountain Terrace Maintenance in Yemen*, by A. Aw-Hassan, M. Alsanabani and A. Bamatraf, July 2000.
 - 04 *Land Tenurial Systems and the Adoption of a Mucuna Planted Fallow in the Derived Savannas of West Africa*, by Victor M. Manyong and Victorin A. Houndékon, July 2000.
 - 05 *Collective Action in Space: Assessing How Collective Action Varies Across an African Landscape*, by Brent M. Swallow, Justine Wangila, Woudyalew Mulatu, Onyango Okello, and Nancy McCarthy, July 2000.
 - 06 *Land Tenure and the Adoption of Agricultural Technology in Haiti*, by Glenn R. Smucker, T. Anderson White, and Michael Bannister, October 2000.
 - 07 *Collective Action in Ant Control*, by Helle Munk Ravnborg, Ana Milena de la Cruz, María Del Pilar Guerrero, and Olaf Westermann, October 2000.
 - 08 *CAPRI Technical Workshop on Watershed Management Institutions: A Summary Paper*, by Anna Knox and Subodh Gupta, October 2000.
 - 09 *The Role of Tenure in the Management of Trees at the Community Level: Theoretical and Empirical Analyses from Uganda and Malawi*, by Frank Place and Keijiro Otsuka November 2000.
 - 10 *Collective Action and the Intensification of Cattle-Feeding Techniques a Village Case Study in Kenya's Coast Province*, by Kimberly Swallow, November 2000.
 - 11 *Collective Action, Property Rights, and Devolution of Natural Resource Management: Exchange of Knowledge and Implications for Policy*, by Anna Knox and Ruth Meinzen-Dick, January 2001.
-

CAPRI WORKING PAPERS

- 12 *Land Dispute Resolution in Mozambique: Evidence and Institutions of Agroforestry Technology Adoption*, by John Unruh, January 2001.
 - 13 *Between Market Failure, Policy Failure, and “Community Failure”: Property Rights, Crop-Livestock Conflicts and the Adoption of Sustainable Land Use Practices in the Dry Area of Sri Lanka*, by Regina Birner and Hasantha Gunaweera, March 2001.
 - 14 *Land Inheritance and Schooling in Matrilineal Societies: Evidence from Sumatra*, by Agnes Quisumbing and Keijuro Otsuka, May 2001.
 - 15 *Tribes, State, and Technology Adoption in Arid Land Management, Syria*, by Rae, J., Arab, G., Nordblom, T., Jani, K., and Gintzburger, G., June 2001.
 - 16 *The Effects of Scales, Flows, and Filters on Property Rights and Collective Action in Watershed Management*, by Brent M. Swallow, Dennis P. Garrity, and Meine van Noordwijk, July 2001.
-