PN-ABY-575
98951

CENTER FOR INSTITUTIONAL REFORM AND THE INFORMAL SECTOR

University of Maryland at College Park

Center Office: IRIS Center, 2105 Morrill Hall, College Park, MD 20742
Telephone (301) 405-3110 • Fax (301) 405-3020

DOES PARTICIPATION IMPROVE PERFORMANCE? ESTABLISHING CAUSALITY WITH SUBJECTIVE DATA

Jonathan Isham, Deepa Narayan and Lant Pritchett

IRIS Reprint No. 67

This publication was made possible through support provided by the U.S. Agency for International Development, under Cooperative Agreement No. DHR-0015-A-00-0031-00 to the Center on Institutional Reform and the Informal Sector (IRIS) and administered by the Office of Economic and Institutional Reform, Center for Economic Growth, Bureau for Global Programs, Field Support and Research.

The views and analyses in the paper do not necessarily reflect the official position of the IRIS Center or the U.S.A.I.D.

Authors: Jonathan Isham, University of Maryland and IRIS Center.
Deepa Narayan, Environment Department, The World Bank.
Lant Pritchett, Policy and Research Department, The World Bank.
Reprinted from *The World Bank Economic Review*, Vol. 9, No. 2, pp. 175-200.

DOES PARTICIPATION IMPROVE PERFORMANCE?

ESTABLISHING CAUSALITY WITH SUBJECTIVE DATA*

Jonathan Isham
IRIS Center,
Department of Economics
University of Maryland, College Park, MD

Deepa Narayan
Environment Department
The World Bank, Washington, D.C

and

Lant Pritchett
Policy and Research Department
The World Bank, Washington, D.C.

*This publication was made possible through support provided by the U.S. Agency for International Development under Cooperative Agreement No. DHR-0015-A-00-0031-00 to the Center on Institutional Reform and the Informal Sector (IRIS) and administered by the Office of Economic and Institutional Reform, Center for Economic Growth, Bureau for Global Programs, Field Support and Research.

THE WORLD BANK ECONOMIC REVIEW, VOL. 9, NO. 2: 175-200

Does Participation Improve Performance? Establishing Causality with Subjective Data

Jonathan Isham, Deepa Narayan, and Lant Pritchett

Data from 121 diverse rural water projects provide strong statistical findings that increasing beneficiary participation directly causes better project outcomes. Three possible econometric objections to these findings are addressed and answered. The subjective nature of the data does not preclude valid, cardinal measures of participation appropriate for statistical analysis. "Halo effects"—changes in the measurement of one variable because of the observed state of another variable—do not seem to induce a strong upward bias in the measurement of participation or project performance. Reverse causation is unlikely: estimation using instrumental variables, data on project timing, and documentation of case studies support the cause-effect relation between participation and better project performance.

An increasing number of development practitioners agree that participation of the intended beneficiaries improves project performance. Participatory development, championed since the early 1970s by mostly noneconomic social scientists and grassroots organizations (Freire 1973; Korten 1980), is increasingly advocated by the largest and most influential aid agencies (UNDP 1993; World Bank 1991). The existence of consensus (or advocacy), however, does not imply the existence of clear and convincing evidence that participation improves performance. Advocates of participation have most often relied on case studies to document the link between participation and performance (Briscoe and de Ferranti 1988; Korten and Siy 1988). These case studies, however, are easily dismissed by skeptics as inconclusive because the small number of cases and the informal method do not allow formal testing of the findings. In response, some studies have used the systematic case study method to establish statistically the relationship between participation and project performance (Esman and Uphoff 1984; Finsterbusch and Van Wicklin III 1987).

Jonathan Isham is with the Department of Economics at the University of Maryland and the center of Institutional Reform and the Informal Sector (IRIS); Deepa Narayan is with the Environment Department at the World Bank; and Lant Pritchett is with the Policy and Research Department at the World Bank. This article was a background paper for *World Development Report 1994*. It is based upon a larger research project initiated and funded by the Participatory Development Learning Group, the UNDP–World Bank Water and Sanitation Program, and the Social Policy and Resettlement Division of the World Bank. The authors would like to thank John Blaxall, Antonio Estache, Peter Lanjouw, and Jim Tybout for helpful suggestions.

© 1995 The International Bank for Reconstruction and Development / THE WORLD BANK

Skeptics have raised three objections to the statistical evidence of a causal impact of participation on project performance. First, participation and project ratings are not objective: subjective data are not appropriate for statistical analysis. Second, the subjectivity in the ranking of projects will lead to halo effects: if investigators believe participation is good, their subjective rankings will overstate the level of participation in highly successful projects and the level of success of highly participatory projects. Third, better project performance may cause increased beneficiary participation rather than vice versa: a mere statistical association is not evidence of a causal impact of participation on performance.

In this article we address and overcome each of the above objections. We start with a brief review of the construction of the data on project performance and beneficiary participation for 121 rural water supply projects and a presentation of the basic statistical results.

In answer to the first objection, we show that the subjective nature of the data does not preclude intersubjectively valid, cardinal measures of participation appropriate for statistical analysis. There is no necessary connection between objectivity and quantification with cardinal numbers. Cardinal measurement using subjective criteria is common. Moreover, the cardinal rankings for each project were created by two different coders, and the degree of intercoder agreement is very high. This agreement indicates that intersubjective reliability can be achieved even for intrinsically subjective concepts such as participation. Finally, we show that assuming cardinality or imposing linearity is not necessary to establish the basic result.

We address the halo effect in coding performance and participation data from project documents in two ways. First, we show that the results are the same if the first coder's performance indicators are regressed on the second coder's participation scores (and vice versa); this test indicates that the coders' subjective ranking does not lead to a halo effect. The primary danger of subjective measurement, however, and one that we cannot address, is having the same individual assess both project success and participation. If that person has strong views about the relationship, these views may induce a bias in the project documents themselves, because the performance of participatory projects (and the participation in successful projects) may be exaggerated. We show, however, that the strength of the performance-participation relationship does not depend on the objectiveness of the success indicator.

The third and most difficult objection to answer is that the existence of an association does not imply causation. Although causality is nearly impossible to establish, we present three arguments in its favor. First, estimation using instrumental variables allows the identification of the impact of exogenous changes in participation, and eliminates the effect on the participation estimates of reverse causation or simultaneity. Second, data on the timing of participation show that participation at early stages improves project performance at every stage, from implementation to maintenance. Third, case studies show that exogenous

changes in participation in ongoing projects have strong impacts on performance.

Even for this limited set of projects, this article is not intended to be comprehensive. We focus on the econometric issues involved in drawing causal inferences from subjective project data. Narayan (1995), drawing on these data (and much more, including many case studies), discusses in greater detail the relationship between performance and participation, the determinants of project success in addition to participation, the underlying determinants of participation, and the mechanisms whereby participation increases overall effectiveness.

I. The Basics

The systematic case review method (Finsterbusch 1990) is used to transform qualitative evaluations in a set of case studies into data suitable for statistical analysis. This method consists of the following basic steps: A conceptual framework for a system of related phenomena, usually in the form of a flow chart, is established, specifying the causal relationships between different subsets of the system; a questionnaire is developed to delineate relevant elements of each of these subsets and to measure their magnitude on the basis of prior knowledge of the system and on the hypotheses to be tested; two independent coders then test this preliminary questionnaire in a small number of case studies by transforming the relevant qualitative information for each question into a cardinal scale; finally, after refining the questionnaire—to match available information and to eliminate redundancy—the coders independently review each of the cases and produce two sets of data that can be evaluated using established statistical techniques.

Data on Rural Water Projects

The data from the 121 rural water supply projects in this study were assembled from project documents, using the systematic case review method (for more on this and similar methodology, see Finsterbusch and Van Wicklin III 1987 and Esman and Uphoff 1984). Ex post project assessments by development agencies typically combine limited quantitative evaluations with subjective judgments of project performance. Each project document was read by two independent readers who coded specific project variables (for example, overall success of project) onto a scale from 1 to 7, creating 144 distinct variables.

Appendix table A-1 shows the list of those coded variables that were used in the analysis along with the basic summary statistics. The variables can be divided into four groups:

- Project performance indicators (for example, overall project effectiveness and the percentage of the water system in good condition)
- Measures of participation (for example, overall beneficiary participation and participation in construction)

4

- Background characteristics or project characteristics that determine project performance (for example, the size of the project and the availability of spare parts)
- Background characteristics or project characteristics that determine participation (for example, the extent to which the agency made participation a goal, and consensus among users about project objectives).

The projects, drawn from forty-nine countries in Africa, Asia, and Latin America, were financed by a range of external donors, including the U.S. Agency for International Development (USAID), the United Nations Development Programme (UNDP), the World Bank, and nongovernmental organizations (NGOs). The range of project costs was from $0.5 million to $250 million, and the water technologies ranged from spring captures to power-pumped piped water systems. The primary objective as well as the common thread of all the projects was water supply, although many of the projects included other components, and the projects were assessed on their overall performance.

The rural water sector was chosen primarily because community participation had became a central feature of the strategy for reaching scattered rural communities with safe drinking water during the International Drinking Water Supply and Sanitation Decade of the 1980s. Given the local common property nature of the good, the difficulty in achieving economies of scale in rural areas, and the need to develop decentralized maintenance systems, local responsibility—generated through participation—is thought to be particularly relevant to the sector.

The participation variable merits some discussion. The measure of participation was not simply a measure of whether potential beneficiaries were surveyed about their preferences. Participation was scored on a continuum, progressing from information sharing, to more in-depth consultation, to shared decisionmaking, to control over decisionmaking. The participation of beneficiaries was considered at three different stages of the project cycle: project design, construction, and operation and maintenance.

The Basic Model and Results

We specify the most general indicators—overall project effectiveness (OPE) and overall beneficiary participation (OBP)—and estimate a simple linear relationship between them. The bivariate relationship between OPE and OBP in the ith project is simply

$$OPE_i = \beta * OBP_i + \varepsilon_{i1}, \tag{1}$$

which is functionally equivalent to the simple correlation of the two variables.[1]

1. In the bivariate linear regression, the regression coefficient is σ_{xy}/σ_x^2 (σ_{xy} is the covariance of y and x), where the correlation coefficient ρ is $\sigma_{xy}/\sigma_x\sigma_y$. The correlation coefficient $\rho = \beta(\sigma_x/\sigma_y)$ can be derived from

The usefulness of this bivariate relationship is limited because other nonparticipation determinants of project performance are excluded. In expanding the model, it proves useful to divide nonparticipation determinants into two groups: those that are fully exogenous and not affected by participation (for example, availability of spare parts), denoted by the matrix Z, and those that are potentially affected by participation (for example, responsiveness of managers), denoted by the matrix W. The multivariate equations are then

$$OPE_i = \beta * OBP_i + \delta_z * Z_i + \delta_w * W_i + \varepsilon_{i2} \tag{2}$$

$$W_i = \gamma * OBP_i + \gamma_z * Z_i + \gamma_x * X_i + \eta_i \tag{3}$$

where X is the matrix of nonparticipation determinants of the Ws. Both Zs and Ws are potential determinants of project performance; Zs, however, represent variables that are not influenced by participation, and Ws represent variables that may be determined (in part) by participation. As indicated in the second equation, Ws may also be determined in part by the Zs and by some other set of variables, Xs. In summary, these multivariate equations state that the performance of a water project depends on beneficiary participation, a set of inputs (Z) not related to participation, and a set of performance determinants (W) that may in turn be determined by participation and other inputs.

The distinction between the Zs and the Ws is important for maintaining the distinction between the partial and the total impact of participation on project performance. In the multivariate regression (equation 2), the β coefficient gives the direct impact of increasing participation, holding all included variables constant:

$$\left.\frac{\partial OPE}{\partial OBP}\right|_{Z=z, W=w} = \beta. \tag{4}$$

But participation may also influence performance indirectly. The total impact of changing participation is the sum of the direct and indirect impacts:

$$\frac{dOPE}{dOBP} = \left.\frac{\partial OPE}{\partial OBP}\right|_{Z=z, W=w} + \frac{\partial OPE}{\partial W} * \frac{\partial W}{\partial OBP} \tag{5}$$

or, in this particular specification,

$$\frac{dOPE}{dOBP} = \beta + \delta_w * \gamma. \tag{6}$$

Thus, the simple partial coefficient with all controls understates the total impact of participation, and the bivariate coefficient (which excludes the Zs and Ws) overstates the impact to the extent to which these determinants are positively correlated with participation.

---

a simple rescaling of either of the possible bivariate linear regression coefficients (for example, y on x or x on y).

Results

Table 1 presents three estimates of the linear association of OBP with OPE: bivariate, limited multivariate (with *Z*s), and full multivariate (with *Z*s and *W*s). For each regression we report the linear regression coefficient on OBP in the OPE regression. In all cases the results are strongly statistically significant and empirically quite large.

The estimated impact of participation on project effectiveness ranges from 0.62 for the bivariate case to 0.24 in the full multivariate case. The multivariate impact is naturally lower than the bivariate effect because of the exclusion of positively correlated nonparticipation performance determinants, as discussed above. Although the sample sizes differ, sample size is not the determinative factor, as shown by the estimates of the bivariate impact in the smaller samples used for the multivariate regressions: (*t*-statistics in parentheses) 0.599 (7.31) for sample size 77, and 0.557 (6.09) for sample size 68. The bivariate effect is an upward-biased estimate of the total effect. The full multivariate estimate of the partial effect is likely biased downward for the total effect. Therefore, the bivariate and full multivariate results create reasonable bounds for the total effect.

How should the coefficients be interpreted? The expected impact of increasing participation from a low level (OBP = 2) to a high level (OBP = 6) is to improve project performance 1 to 2.5 points (on a scale of 1 to 7). A one-standard-deviation increase in participation (in appendix table A-1, the standard deviation is 1.7) is associated with an increase in performance of between 0.41 points (for the full multivariate model) and 1.05 points (for the bivariate model). The interesting—and intuitively appealing—results of all the regressions (limited and full multivariate) are reported in appendix table A-2 and are discussed in Narayan (1995).

Table 1. *Basic Estimation Results from Regressing Overall Project Effectiveness on Overall Beneficiary Participation*

Item	*Bivariate model*	*Limited multivariate model*[a]	*Full multivariate model*[b]
Coefficient	0.62	0.28	0.24
	(10.6)	(5.3)	(3.8)
	<11.4>	<5.3>	<3.1>
Sample size	121	77	68
R^2	0.49	0.86	0.89

Note: Values are from ordinary least squares (OLS) regressions in which the dependent variable is overall project effectiveness. The regressions include constants for which results are not reported. *t*-statistics are in parentheses. White heteroskedasticity-consistent *t*-statistics are in angle brackets.

a. The limited multivariate model includes nonparticipation determinants of project performance that are fully exogenous, the *Z* variables. Regression results for the *Z* variables are reported in table A-2.

b. The full multivariate model includes the *W* variables, nonparticipation determinants of project performance that are potentially affected by participation, and the *Z* variables. Regression results for the *Z* and *W* variables are reported in table A-2.

Source: Authors' calculations.

We note here that the choice of the *Z*s and *W*s in this study was neither entirely straightforward nor driven by rules. However, all the results have proved robust to a number of variations of the model, and we feel that the choice of control variables is not of primary interest. Having 144 coded variables with only 121 projects meant that mechanical procedures for selecting variables would lack degrees of freedom and would not likely be of much help. Moreover, many of the variables were clearly overlapping and likely to be collinear. After some experimentation, we based variable inclusion on three criteria: decent intercoder reliability, prior judgments about the best choice among collinear sets of variables, and impact on the estimate of participation (we never dropped any variable that seriously affected the estimate of the participation coefficient). In none of the experiments were the results on participation substantially different from the full multivariate case reported in table 1.

We were more than generous in our inclusion of potential performance factors, including a total of eighteen nonparticipation variables. The participation variable thus easily passes this kitchen-sink test of throwing all plausible variables into the regression. The danger of inadequate controls for other determinants of project performance is not nearly as serious as the three problems we discuss in the following sections.

Heteroskedasticity, a typical econometric problem that receives a fair bit of attention, is not a problem with the present results for two reasons. First, the White heteroskedasticity-consistent standard errors are roughly the same as those obtained using ordinary least squares (OLS). For the bivariate case, the *t*-statistics are 11.4 with White compared with 10.6 with OLS; for the full multivariate case, they are 3.1 with White and 3.8 with OLS. Second, in scoring variables, the coders recorded their subjective assessment of the reliability of the score assigned. When these reliability measures were used to weight observations, the results were roughly the same.

II. Subjective Cardinal Data

The first objection to studies—and results—of this kind is that the data generated by the systematic case review method are subjective. According to this skeptical view, subjective data are unreliable, ordinal, or both, and therefore inappropriate for statistical analysis. In this section we show that our data are subjective, yet reliable and cardinal. First, we argue that the automatic association of subjective phenomena with ordinal data is incorrect. Second, the degree of intercoder agreement on the scoring of the major variables reveals that the subjective measurement error, although present, is a minor source of variation. Third, using techniques appropriate for ordinal data does not dramatically change the results, and the constraints imposed by linear regression analysis are also not rejected by the data.

8

Subjective and Objective, Ordinal and Cardinal

Economic theory often creates a presumption that objective phenomena (such as numbers of oranges or relative prices) have a natural cardinal metric (such as integers or real numbers) whereas intrinsically subjective phenomena (such as consumer utility) allow only ordinal comparisons that are better or worse, especially intersubjectively. Ordinality stems directly from the basic theory of mapping a binary preference relation into a utility index. With only the barest restrictions imposed on the preference relation (complete, reflexive, and transitive), a numerical utility index can be derived, but any monotone transform of that index represents the same preferences equally well. This is not to say that all utility functions are ordinal. Often additional assumptions are imposed that imply cardinal functions that are unique only up to an affine transformation (for example, von Neumann–Morgenstern utility functions).

This distinction between cardinal and ordinal is critical for empirical work. Although both cardinal and ordinal data can rank phenomena, only cardinal data can be used to tabulate numbers and to directly compare values of the phenomena being measured. Common statistical techniques such as correlations or linear regression cannot be applied to ordinal data. For instance, if x were an ordinal measure of participation, estimating the linear model $y = \beta * x$ could produce different results than estimating $y = \beta * f(x)$, where $f(x)$ is a monotone transform of x, even though x and $f(x)$ would represent exactly the same information. Therefore, any statistical procedure that relied on summing observations (or any other comparison of the magnitude of the distance between observations) would be invalid for ordinal data.

Nevertheless, the data used in this analysis, created by applying the systematic case review method, are subjective, yet cardinal. Our data on rural water projects are doubly subjective: the original project evaluator subjectively assessed and described the amount of participation in each project, and a coder later read the evaluator's report and subjectively assigned a level of participation to that project. If this process generated ordinal data, empirical analysis would be difficult. But note that in everyday life we observe many events that generate subjective, cardinal data Contests are the most obvious example. When hogs, figure skaters, or bodybuilders compete, judges assign cardinal scores to subjective criteria: quality of coat for hogs; artistic impression for figure skaters; and muscle tone for bodybuilders. Grades for academic papers are another familiar example: a professor's subjective evaluation of a humanities paper is given a cardinal score. In each case, these subjectively assigned scores are added, averaged, and tabulated in ways only appropriate to cardinal data. Of course, the judging and grading criteria are created to achieve intersubjective consensus. This means that judging requires training to achieve this level of intersubjective agreement. For instance, judges of livestock contests are occasionally judged on the degree to which their subjective judgments conform to those of established judges (Herren 1984). Thus, the question for this data set on characteristics of

water projects is not whether the data are subjective, but whether the cardinal scores based on them are reliable. The notions of reliability and validity play a large role in the literature on educational and psychometric testing. Reliability typically refers to whether different versions of the same test on the same individual will produce the same result (for example, whether repeated I.Q. tests will produce the same result). Validity refers to the usefulness of the tests in some application (for example, the usefulness of scores on the Graduate Record Exam in identifying successful graduate students). Although we now show reliability, we have no external check on validity.

Intersubjective Agreement

Because project variables were scored from the same documents by two independent coders, the coherence of their separate scores illuminates the overall reliability of the variables. Table 2 presents two measures of the crosscoder agreement. The correlations between the scores of coder A and coder B are strikingly high: 0.95 for OPE and 0.92 for OBP. The average absolute value of the difference in the scores (on a scale of 1 to 7) is 0.36 for OPE and 0.55 for OBP. The difference is quite small: most scores either agree or differ by just one point. For each of the two major variables, the coders disagreed by two or more points on only one project.

This high degree of intersubjective consensus between reasonably independent coders has two important implications: it implies that the characteristics of the project could be reliably gauged from the project documents (although the reliability was much lower for some other variables in the data set) and places a relatively tight bound on the magnitude of measurement error. A correlation coefficient of 0.9 implies that the noise from measurement error is roughly 10 percent of the variance of the observed variable.[2]

Table 2. *Crosscoder Reliability*

Variable	*Correlation between coders A and B*	*Average absolute difference in scores*[a]
Overall project effectiveness (OPE)	0.95	0.36
Overall beneficiary participation (OBP)	0.92	0.55

a. The average absolute difference between the scores of the two coders. The scores are on a scale of 1 to 7.

Source: Authors' calculations.

2. If two observations differ by only measurement error, then the correlation between the two observations is

$$\rho = \sigma^2_{x^*} / \left(\sqrt{\sigma^2_{x^*} + \sigma^2_{\varepsilon_A}} \sqrt{\sigma^2_{x^*} + \sigma^2_{\varepsilon_B}} \right)$$

where $\sigma^2_{x^*}$ is the variance of the true variable and $\sigma^2_{\varepsilon_{A(B)}}$ is the measurement error variance for coder A(B). If the measurement error variance for both coders is equal $\sigma^2_{\varepsilon_A} = \sigma^2_{\varepsilon_B} = \sigma^2_{\varepsilon}$ a correlation of 0.9 implies that the ratio of measurement error to true variance $\sigma^2_{\varepsilon} / \sigma^2_{x^*}$ is about 0.1.

Testing Linearity or Cardinality

We examine in several ways whether the assumption of cardinality affects the results of the analysis of the relation between project performance and participation. We do a simple test for linearity of the relationship. We estimate the participation effect by treating the participation data as if they were ordinal, using dummy variables for each level of participation. We estimate the relationship by using the ordered probit estimation technique, which treats the project performance data as ordinal. Finally, we estimate the model with both participation and performance treated as binary variables. Of course, these techniques do not prove cardinality of the data; they do, however, show that the basic results on beneficiary participation are unaffected either by allowing for the possibility that the data are ordinal or by our assumptions of functional form.

The first approach argues that if the participation data were in fact ordinal, the relationship between OPE and OBP would not be linear. That is, the true underlying relationship between the ordinal variables would not be invariant with respect to arbitrary transformations (for example, squaring) of the data. The second column of table 3 presents a test of linearity allowing for a slope shift depending on the value of participation. When participation is low (OBP less than 3.5), the slope is β_1, and when participation is high (OBP greater than 3.5), the slope is $\beta_1 + \beta_2$. The estimates suggest that the incremental impact of participation is larger at higher levels (the slope is 0.466 for OBP less than 3.5 compared with 0.781 for OBP greater than 3.5). This difference is not statistically significant; a test of the difference resulted in a low t-statistic and a declining adjusted R^2.

The second approach treats participation as if it were ordinal while treating performance as cardinal. Each discrete level of the participation variable is entered into the performance equation as a dummy variable. The first dummy equals 1 if OBP is less than or equal to 1.5; 0 otherwise. The second dummy equals 1 if OBP is greater than 1.5 and less than or equal to 2.5. And so on. Ranges for the variables were specified to generate these dummy variables because the averages for the coders' scores were not always integers. This functional form imposes no a priori constraints on the effect of the independent variable. The results in the third column of table 3 show a strong participation effect—performance increases for each performance category—without any strong indications that this statistically unconstrained fit is tremendously superior to the imposition of linearity. The adjusted R^2 is lower with the series of dummy variables, but this combines the effects of relaxing linearity with imposing a discrete step function. The implied slope from category to category (from the differences in the coefficients and means of participation) ranges from 0.46 to 0.87. The mean is 0.642, roughly equivalent to the overall linear slope of 0.623. Increases in participation have roughly the same impact along the range from low to high participation.

The third approach uses ordered probit estimation by creating a categorical variable for each range of performance. Again, because the averages of the

Table 3. *Estimation Results for Alternative Functional Forms*

Variable	Linear	Linear with kink	Binary participation measure: Coefficient	Binary participation measure: Implied slope[a]	Ordered probit[b]	Binary performance and participation measures[c]
Participation coefficient	0.623 (10.6)	0.466 (3.007)			0.552 (7.906)	0.552 (7.34)
Participation coefficient after the kink point[d]		0.315 (1.33)				
Constant	1.79 (7.52)	2.09 (6.13)				
Second constant		0.027 (0.066)				
Cutoff points for ordered probit						
First					0.957	
Second					1.39	
Third					2.16	
Fourth					2.93	
Participation dummy variable						
≤1.5			2.55			
1.5< · ≤2.5			3.06	0.460		
2.5< · ≤3.5			3.59	0.624		
3.5< · ≤6.5			4.25	0.609		
4.5< · ≤5.5			5.16	0.877		
>5.5			5.74	0.639		
$\bar{R}^2$[e]	0.481	0.480	0.459		0.1828	0.306

Note: Estimates are based on ordinary least squares (OLS) regression unless otherwise noted. *t*-statistics are in parentheses.

a. The implied slope is the difference in the coefficient across participation categories divided by the difference in the means of the participation variable across categories.

b. The magnitude of the coefficient in the ordered probit model is not directly comparable to the OLS models or the binary model in the sixth column. The fact that both values are 0.552 is a pure coincidence.

c. The binary model is estimated as a linear probability model.

d. For values of the participation variable greater than 3.5.

e. The $\bar{R}^2$ is not comparable between the linear regressions and the ordered probit pseudo AR^2 or the binary model because the dependent variable is transformed.

Source: Authors' calculations.

coders' scores are used, the variables are not just levels. Five categories were created by dividing at OPE levels of 2.5, 3.5, 4.5, and 5.5. Ordered probit estimation uses only the information that performance categories are different and that higher performance categories represent better levels of performance. It does not use the magnitude of the differences and hence does not assume that the performance variable is cardinal. The fifth column of table 3 reports the ordered probit results, which again show a strongly significant effect of participation. This comes through in the multivariate ordered probit results as well. The slope coefficient in an ordered probit estimation is not the marginal effect on the probabilities. Calculations from the results, however, imply that an increase of

12

participation from 3.5 to 4.5 would reduce the probability of failure (OBP falling in the lowest performance category) by 62 percent (from 0.164 to 0.063) and increase the probability of excellent performance (being in the highest performance category) by more than 100 percent (from 0.157 to 0.325).[3] Again, these results are broadly consistent with the results of simple linear OLS.[4]

The final approach, which checks on cardinality, treats both the performance and the participation data as ordinal. For both project performance and participation, a binary variable takes a value of 1 if the score is high (OBP greater than 3.5) and 0 otherwise. This procedure is valid even if the data are ordinal; binary variables would be unaffected by monotone transforms. If the data are in fact cardinal, however, this procedure is very inefficient because it throws away all of the information about variation within each of the two performance categories. The final column of table 3 reports the results of this linear probability regression. The performance-participation effect remains evident with this crude transformation of the data.

The subjective nature of the data per se appears to have no impact on the results. High intersubjective reliability of the measures was achieved. The results appear to be broadly consistent with a simple linear model, and treating either performance or participation data as if they were ordinal produces similar results.

III. Halo Effects

A potentially more serious problem than the intrinsically subjective nature of the data is that either the initial evaluator of the projects or the coders themselves succumbed to the plausible assumption that all good things go together: the halo effect. This psychological tendency to associate all good things has been discussed in a number of fields. In particular, there is a large literature in human resource management about the halo effect problem in assessing performance.

3. The formulation for calculating the incremental change in probability of observing the dependent variable in one of J categories (where the J categories are defined by whether they fall between endogenously determined cutoff points: $0 < \mu_1 < \mu_2 \ldots < \mu_{J-1}$) with respect to a change in a dependent variable in an ordered probit model is $\partial Prob[y = 0]/\partial X = -\Phi(\beta' X)\beta$ for the lowest category. For the highest category the formula is $\partial Prob[y = J]/\partial X = \Phi(\mu_{J-1} - \beta' X)\beta$, where $\Phi(.)$ is the value of the standard normal probability density function (Greene 1990). Using these formulas at OBP = 3.5, the impact of changing OBP for the probability of OPE's occurring in the lowest category is −0.034, and for the highest category the marginal impact of OBP on the probability is 0.132.

4. Using the simple OLS model, we can calculate the change in probability of project failure as the change in the probability that OPE is less than 2.5 when OPE equals 3.5 compared with the probability that OPE is less than 2.5 when OPE equals 4.5 (which is not exactly comparable as the cutoff points are endogenously estimated in an ordered probit). The first probability for instance would be $Pr(\alpha + \beta 3.5 + \varepsilon < 2.5)$, which, given our estimates of $\alpha = 1.79$, $\beta = 0.623$, and $\sigma_\varepsilon^2 = 1.246$ and assuming that the error term is normal, is the same as the probability that x is less than −1.316, where z is a standard normal. A similar calculation can be performed for OPE greater than 5.5. With our linear estimates the probability of OPE less than 2.5 falls by 0.064 (from 0.094 to 0.030) and the probability of OPE greater than 5.5 rises by 0.123 (from 0.085 to 0.208).

Outstanding performance in one dimension or characteristic (even a potentially irrelevant characteristic, such as physical attractiveness) may tend to bias upward the evaluation of other dimensions or characteristics. Hammermesh and Biddle (1994) find that plain people make about 5 percent less and attractive people 5 percent more than persons of average attractiveness. However, for a recent dissenting view on the importance of halo effects in performance evaluation, see Murphy, Jako, and Anhalt (1993).

The halo effect occurs when the measurement of one variable is affected by the observed state of another variable. This systematic measurement error will induce an association between two subjectively measured variables even in the absence of any true relation between the underlying variables. In our study, the halo effect may occur at two stages. The evaluators may have falsely attributed participation to successful projects (or vice versa), or the coders—searching the project documents for evidence of project participation—may have been affected by their simultaneous assessment of project success despite their efforts to remain objective. The second possibility is particularly dangerous. In this study the two coders knew the purpose of the empirical exercise and may have had some strong prior beliefs as to the expected outcome.[5]

There is nothing we can do about the potential halo effect of the original evaluations. We know that the project reports were regular parts of the institutional evaluation cycle and that it is doubtful that the financing agencies had a particular stake in promoting participation. It can also be expected that the many individuals writing the project documents would have widely different beliefs about the importance of participation, so that a uniform bias in the firsthand assessments would be unlikely.

We explore three methods of addressing the problem of the potentially serious halo effect in the coding process. Note that the results in tables 1 and 3 are based on the average of the two coders' assessments. In the first method we estimated the same models using only data from coder A and from coder B. Differences in these two assessments may reflect differences in the halo effect between the coders. In the second, we estimated the same models using coder A's assessment of the explanatory variables (including OBP) with coder B's assessment of the dependent variable (OPE). Because coder A's assessments of participation and the other potential determinants are not affected by coder B's performance assessment, the halo effect bias should be reduced (although the confounding effect of pure measurement error in the coders' assessments will be important). In the third method, we use project performance indicators—created by the coders—that, by their nature, are more objective than others. If halo effects were present, they would be more likely to appear for the more subjective indicators.

5. In fact, one of the coders had participated in a previous similar empirical study that had found significant effects of participation. The other coder was a graduate student who was hired and trained to code for this exercise but was new to the field and to the topic.

Table 4. *Estimation Results by and across Coders*

	Scores used				
Model	*Average for coders A and B*	*Coder A only*	*Coder A's OPE and coder B's OBP*	*Coder B only*	*Coder B's OPE and coder A's OBP*
Bivariate model					
Coefficient	0.62	0.60	0.62	0.60	0.57
	(10.6)	(10.1)	(10.3)	(9.7)	(9.3)
Sample size	121	111	116	111	116
R^2	0.49	0.49	0.48	0.46	0.43
Full multivariate model					
Coefficient	0.24	0.23	0.26	0.21	0.25
	(3.8)	(2.6)	(2.1)	(2.0)	(2.7)
Sample size	68	37	46	46	37
R^2	0.89	0.94	0.85	0.89	0.94

Note: Results are presented for OLS regression of overall project effectiveness (OPE) on overall beneficiary participation (OBP). *t*-statistics are in parentheses.
Source: Authors' calculations.

Results by Coder and across Coders

Table 4 shows the results of OLS estimation using the average scores of the two coders (A and B) and using each coder's scores. The table also shows the results of regressing coder A's score for OPE on coder B's score for OBP and of regressing coder B's score for OPE on coder A's score for OBP. The differences for both the bivariate and multivariate models are very small. In both models the coefficient does not systematically change, whether we use the average of the coders' scores, each coder's own scores, or one coder's dependent variable scores on the other coder's independent variable scores.

How reassuring are these crosscoder results? Suppose that A's observation on project performance is the truth (OPE*) plus some random noise (ε_A), plus an upward bias based on A's observation of participation. Then

$$OPE_A = OPE^* + \delta_A * OBP_A + \varepsilon_A. \tag{7}$$

Coder A's observation on participation is just the truth plus random error:

$$OBP_A = OBP^* + \eta_A. \tag{8}$$

In this case, if we assume that the pure observational errors are uncorrelated, the coefficient of regressing performance on participation will still be biased upward by the halo effect. If the true structural relationship were

$$OPE = \beta * OBP + \varepsilon, \tag{9}$$

the estimated coefficient would be

$$\beta + \delta_A. \tag{10}$$

So even if there were no structural relationship between the true variables ($\beta = 0$), the estimate of the participation effect could be spuriously positive because of halo effects.

Table 5. *Results of Monte Carlo Simulations of the Combined Effects of Measurement Error and Halo Effects Using Crosscoder Information*

	Degree of measurement error (k)					
	Zero, k = 0		*Low, k = 0.25*		*High, k = 0.5*	
Degree of halo effect (δ)	*Coder A's scores only*	*Coder A's OPE on coder B's OBP*	*Coder A's scores only*	*Coder A's OPE on coder B's OBP*	*Coder A's scores only*	*Coder A's OPE on coder B's OBP*
Zero, $\delta = 0$	0.5	0.5	0.40	0.40	0.33	0.33
Low, $\delta = 0.25$	0.75	0.75	0.65	0.60	0.58	0.50
Moderate, $\delta = 0.5$	1	1	0.90	0.80	0.83	0.67
High, $\delta = 1$	1.5	1.5	1.4	1.2	1.3	1.0

Note: The results are the average estimates from 1,000 replications of each of the 120 observations of the model: $y = \beta x + \varepsilon$, $\beta = 0.5$, x, $\varepsilon \sim N(0,1)$. The values in the first two columns of the first row are $\beta = 0.5$, the true values with no measurement error and no halo effect. A(B)'s observations on the x variable are subject to measurement error of the form:

$$x_{A(B)} = x^* + \sqrt{k}\eta_{A(B)},$$

x^* is the true value, and $\eta \sim N(0, 1)$ where A(B) indicates that A and B have independent random measurement error of proportion k. The observations on the dependent variable y are determined by

$$y_{A(B)} = y^* + \delta x_{A(B)}$$

so that the measurement error of A or B influences the measurement of y by a common "halo" factor of δ.
Source: Authors' calculations.

Given this background, why does having another coder matter? If the observations—and scoring—of participation are completely objective, using a second coder's data will have no effect: A's and B's observations on participation will be identical ($\text{OBP}_A = \text{OBP}_B$). And if the degree of halo effect is similar ($\delta_A = \delta_B$), the bias on performance will be equivalent. If the observations of participation are subjective, then the halo effect bias should be less using crosscoder data because the pure subjective component of B's assessment does not affect the bias in the performance measure. However, to the extent the measure of performance is truly subjective, this measurement error argues against the intersubjective reliability above and induces downward bias in the estimates caused by classical measurement error.

Table 5 shows Monte Carlo simulations of the combined effects of the halo effect bias and of pure measurement error, using different assumptions about the relative strengths of the two effects. Unfortunately, the simulations show that both underestimation and overestimation of the true coefficient are possible when coders' scores are crossed, depending on the ratio of the halo and measurement error effects. With high measurement error (fifth and sixth columns), crossing the coder rankings should produce lower estimates than using the rankings of a single coder for all possible strengths of halo effect. For inputs and outputs with no measurement error (the first and second columns) and high degrees of the halo effect, crossing the coder rankings does not help because it produces the same estimates (with a similar upward bias).

Evidence of the relatively high intersubjective reliability (as well as the instrumental variable results below) suggests low but nonzero measurement error. The

Table 6. *The Impact of Participation on Various Indicators of Performance*

Model	Overall project effectiveness	Objective value of benefits	Percentage of water system in good condition	Percentage of target population reached
Bivariate model				
Coefficient on participation	0.62 (10.6)	0.53 (10.3)	0.54 (6.4)	0.29 (5.30)
Sample size	121	120	98	118
R^2	0.49	0.47	0.29	0.19
Full multivariate model				
Coefficient on participation	0.24 (3.8)	0.27 (.36)	0.29 (2.4)	0.25 (2.50)
Sample size	68	68	60	68
R^2	0.89	0.79	0.77	0.47

Note: Each column presents the results of OLS regression of the bivariate and full multivariate models using different objective indicators of project success. *t*-statistics are in parentheses.
Source: Authors' calculations.

ratio of the measurement error variance (σ_ε^2) to the true variance ($\sigma_{x^*}^2$), that is, $\sigma_\varepsilon^2/\sigma_{x^*}^2$, is between 0.1 and 0.25. In this range of measurement error (third and forth columns), when the coders' OPE and OBP rankings are crossed there should be a modest but significant change in estimates if the halo effect is strong. The lack of a consistent downward movement of the estimates (in the multivariate case they actually rise) suggests at least that the halo effect is not dominating the results.

Results by Performance Indicator

The third method of evaluating the halo effect is to examine the impact of participation on project performance indicators that vary in their objectivity. In addition to the omnibus measure OPE, several more objective indicators of project success were coded, including the percentage of the water system in good condition and the percentage of the population target reached. To the extent that these more objective phenomena are relatively less susceptible to halo overestimation, the possible halo effect bias should be reduced.

If the true coefficient is equal across models using different dependent variables (which is not clear—see table 6), the more objective indicator should be systematically lower than the upwardly biased subjective indicator. Table 6 presents these results. There is no evidence that the more subjective indicators (such as OPE) have systematically larger estimated impacts.

IV. Joint Determination and Causality

The previous two sections have answered possible skepticism about the strong statistical association between performance and participation. Other skeptics may accept the statistical association between participation and performance but deny that this association reveals cause and effect. In this view, the data do not

show that greater participation causes better project performance but simply that participation and performance happen to be related. Indeed, there are at least two good reasons to believe that an association between participation and performance may not be causal.

First, there could be reverse causation: projects that are exogenously better might induce greater beneficiary participation. Reverse causation makes sense, especially when performance and participation depend on a sequence of actions. Once it is clear that a project is failing, potential beneficiaries may be less likely to participate because they perceive a relatively low benefit from their own participation, which is unlikely to alter the project outcome.

Second, joint determination of project success and beneficiary participation may be driven by a third local or project attribute. For example, if dynamic leaders induce both project performance and participation, performance and participation data will be strongly associated with each other—even without an independent causal effect of participation on performance. Although we have tried to address this concern over spurious correlation with the inclusion of possible performance determinants, it would not take too clever a skeptic to name a large list of excluded variables (some of which are unobservable even in principle) that could affect both performance and participation. We use three approaches to resolve the problems of reverse causation and joint association and to demonstrate a causal impact: estimation with instrumental variables, timing, and case studies.

Instrumental Variables

One econometric solution to the problem of identifying a structural relationship is estimation with instrumental variables (IV). This estimation avoids the problem of the joint determination of the independent and dependent variables: in estimating the coefficients, it eliminates that part of the variation in the independent variable that is caused by variation in the dependent variable. The vehicle for eliminating that variation is a third variable (the instrument), which affects only the independent variable and not the dependent variable.

Estimation with instrumental variables requires a variable that affects participation but that neither affects performance directly, nor is affected by it. This variable is used as the instrument for purging the participation variable of any performance-related component. When the participation effect is estimated using only the part of participation variation that is correlated with the variation in the instrument, the resulting estimate is free of reverse causation. Because better performance does not affect the instrument, the reverse effect of better performance on participation is eliminated and cannot bias the results.

For the following model

$$OPE = \beta * OBP + \delta_z * Z_i + \delta_w * W_i + \delta_v * V_i + \varepsilon_i \tag{11}$$

$$OBP = \alpha_z * Z_i + \alpha_w * W_i + \alpha_v * V_i + \eta_i \tag{12}$$

Table 7. *Comparison of OLS and Instrumental Variables Estimates of the Participation-Performance Relation*

		Instrumental variables					
Model	*OLS*	*Extent to which participation was a goal*	*Percentage of investment costs paid by users*	*Net benefits of participation*	*Organization based on local collectives*	*Prior commitment of clients*	*All instruments*
Bivariate model							
Coefficient	0.63	0.70	0.59	0.77	0.74	0.97	0.86
	(10.6)	(10.2)	(7.3)	(10.6)	(6.3)	(7.54)	(10.4)
Sample size	120	120	113	120	98[b]	105	90
R^2	0.488	0.482	0.476	0.453	0.507[b]	0.378	0.521
First-stage R^2 [a]	n.a.	0.763	0.573	0.701	0.364	0.326	0.816
Limited multivariate model							
Coefficient	0.28	0.34	0.32	0.36	0.15	0.39	0.37
	(5.25)	(5.2)	(3.6)	(5.4)	(1.28)	(3.00)	(3.57)
Sample size	77	77	75	77	66	72	63
R^2	0.862	0.860	0.861	0.858	0.855	0.863	0.865
First-stage R^2 [a]	0.401[c]	0.826	0.643	0.803	0.719	0.559	0.857

n.a. Not applicable.

a. Unadjusted R^2 of the first-stage regression of participation on the instruments (which in the multivariate case includes all variables in the performance equation).

b. Because the sample sizes are not the same, the results are not strictly comparable in all columns. In particular, the IV R^2 values are less than the OLS R^2 when run for the same sample.

c. This is the R^2 of participation regressed on all the Z variables that are included in the performance equation. The increment to the R^2 for each instrument can be calculated as the difference with this column.

Source: Authors' calculations.

all *V*s that are included in the participation equation ($\alpha_v \neq 0$) but excluded from the performance equation ($\delta_v = 0$) are legitimate instruments. The *V*s provide a source of variation in participation that is exogenous to performance. Neither the *Z*s nor the *W*s are valid instruments, because they directly affect both participation and performance.

To choose appropriate instruments, we need a positive model for participation. Hypotheses based on the larger study of these water projects (Narayan 1995) and on other statistical work (Basu and Pritchett 1994), as well as theoretical literature on the determinants of participation, generated a set of equations for estimating the effects of participation (appendix table A-3 shows the full first-stage regressions). We identify five variables as legitimate instruments: the extent to which participation was a project goal, the percentage of investment costs paid by users, beneficiaries' overall net benefits of participation, the extent to which organization was based on local collectives, and prior commitment of clients. We hypothesize that each of these phenomena may directly affect participation but should have no independent, direct effect on project performance after controlling for participation.

In table 7 the OLS and IV results are compared—for both the bivariate and the limited multivariate case. The full multivariate case is excluded because it loses too many degrees of freedom; although the results are empirically similar, they are less precise. The estimated impact of participation increases with IV estimates. For instance, when the extent to which participation was a project goal is used as an instrument, the bivariate impact rises from 0.63 to 0.70, and the multivariate rises from 0.28 to 0.34. The IV estimation produces a higher and statistically significant estimate for each of the instruments used individually. The only exceptions are for the bivariate case in which percentage of investment cost paid by users is used as an investment and for the multivariate case in which organization based on local collectives is used. In the second case the coefficient drops to 0.15 and is statistically insignificant, probably because of the low power of the instrument. When all instruments are used together, the impact rises to 0.86 in the bivariate case and to 0.37 in the limited multivariate case.

What do these IV estimates tell us? The basic statistical relationship—high correlation between participation and performance—would also occur if better project performance caused greater participation. As clean water is delivered in the early stages of a project, more potential beneficiaries may want to get involved. But if this were the causal story, the IV technique would cause the estimates to fall by removing this upward simultaneity bias. The rising coefficients reported in table 7 are consistent with causality running from higher participation to better project performance in the presence of some measurement error.

The IV results allow us to compute an independent estimate of the magnitude of pure measurement error. Even if the OLS estimator is inconsistent with measurement error, the IV estimator is consistent, and the ratio of the estimates converges to

$$plim(\hat{\beta}_{OLS}/\hat{\beta}_{IV}) = \sigma^2_{x*}/(\sigma^2_{x*} + \sigma^2_{\varepsilon}).$$

With the reported estimates, this ratio is between 0.8 and 0.9; for example, 0.63/0.70 (bivariate) or 0.28/0.34 (limited multivariate). The ratio of the variance with measurement error to the true variance, $\sigma_\varepsilon^2/\sigma_{x^*}^2$, is between 0.11 and 0.25. This is consistent with (although somewhat higher than) the estimates of measurement error from the correlations of crosscoder reliability in table 2.

In using this technique, it is useful to test whether the assumptions made in obtaining the IV results are valid. Note that because one variable—beneficiary participation—may be endogenous, at least one instrument must be used to identify the model.[6] But if it can be unambiguously accepted that one instrument directly affects only the independent variable, then the validity of any other instrument can be tested. Indeed, the entire set of instruments can be tested.

We believe that the extent to which participation was a goal is the most plausibly exogenous variable among the individual instruments. There is no reason to believe that participation as a goal would by itself lead to better performance, except insofar as it actually raised participation. When each of the other instruments is tested, conditional on the validity of this variable (using a Hausman-Taylor test), we fail to reject the exogeneity of the other instruments in every case. When the entire set of instruments is tested, we do not reject the validity of the instrument set in either the bivariate or multivariate case. The value of the Sargan test with the full set of instruments is 7.03 (significance level 0.133) in the bivariate and 5.24 (significance level 0.263) in the multivariate estimates. Our set of instruments stands up to the available tests for instrument validity. Of course, the major objection to these tests is that they tend to be of very low statistical power (that is, these tests will often fail to reject a hypothesis that is false). Therefore a failure to reject the instruments cannot be taken as compelling evidence for accepting the instruments.

Timing

Evidence on causality also can be observed from the timing of the project cycle. If the association between participation and project performance was not causal, we would see no association between events that occur before project completion—proximate determinants of project performance—and beneficiary participation. Table 8 reports the impact of participation on quality of implementation, effectiveness of operations and maintenance, and maintenance after one year. We find that in all but one (multivariate) case, beneficiary participation is a statistically significant input of these proximate determinants. If project effectiveness was causing participation rather than vice versa, we would not expect to see this result. Roughly these same results also hold true when, rather

6. Heuristically, the problem is that we need to test the exclusion of the instrument from the performance equation. However, we cannot test the exclusion directly (say by a t-test of the inclusion of the instrument) because in the presence of endogeneity the coefficient on the potentially endogenous variable is inconsistent when it is not instrumented and hence the t-test on the instrument would be biased.

Table 8. *The Impact of Beneficiary Participation on the Proximate Determinants of Project Performance*

Independent variable	*Bivariate model*	*Limited multivariate model*	*Full multivariate model*
Quality of implementation	0.53	0.17	0.21
	(9.3)	(2.7)	(2.7)
Effectiveness of operation and maintenance	0.49	0.14	0.11
	(7.4)	(2.0)	(1.1)
Maintenance after one year	0.43	0.16	0.18
	(6.6)	(2.0)	(1.8)

Note: Values are OLS estimates. *t*-statistics are in parentheses.
Source: Authors' calculations.

than overall participation, the participation at various stages (which therefore precede outcomes) is used as the independent variables.

Case Studies

Case studies of individual projects also help to resolve questions about causality between beneficiary participation and project effectiveness. Narayan (1995) documents the specific effects of participation on performance in many of the projects included in this study. Two of these case studies—in which exogenous shifts in participation during implementation changed project outcomes—clearly illustrate the direction of causality that we have explored econometrically above.

Phase I of the Bank's Aguthi Rural Water Supply Project in Kenya was conducted without community participation. The project, which involved piped water systems, was so plagued with problems—construction delays, cost overruns, and disagreements over consumer payment methods—that it came to a standstill. At this point, there was a substantial shift in the level of beneficiary participation. The project was redesigned, and local leaders organized themselves into the Aguthi Water Committee. Working with project staff, they mobilized community support for the project. After public stakeholder conferences, community members began to contribute labor and finances. Phase II of the project was completed on schedule and within budget. The communities continued to pay monthly tariffs for the new water service, and operation and maintenance of the system were handled successfully, in cooperation with the government parastatal.

The goal of the Wanita, Air dan Sanitasi (WAS) program in Nusa Tengarra Timur, Indonesia, was to help community groups launch and manage their own water system. A water group in the village of Silla was formed in 1986 as WAS began. Initially, the group waited for the arrival of a government team to dig a borehole, but none came. When the group members realized that they could not rely on immediate government assistance, they increased their level of participation in the project. They negotiated water rights with a neighboring water group, collected building materials, and built three water tanks, with only a small

amount of outside technical assistance. By 1988 a new well was under construction, financed by their own contributions. This increased level of participation was maintained and led to the project's sustained success.

V. Conclusion

We began by showing the existence in project-level data of a strong association between project performance and beneficiary participation. We then addressed and answered the three econometric objections to acknowledging this association. The subjectivity of the data is not an overwhelming problem. The halo effect does not appear to induce a strong upward bias. Most important, strong arguments support the cause-effect relation between participation and project performance. This article, together with the more comprehensive work of Narayan (1995) in particular, provides development practitioners—including early and recent converts to the participatory approach—with strong statistical findings that increasing participation directly causes better project outcomes.

Four questions that are important for practice and policy are not explored here. First, does participation directly cause better project performance across all sectors? We cannot blindly apply the results in this study across all sectors, because these data are limited to rural water supply projects. The economic characteristics of rural water as a good would seem to promote the importance of direct beneficiary participation; these economic characteristics vary across goods provided by projects in other sectors.

Second, what policy instruments help to achieve more effective participation? Project beneficiaries, staff in project agencies, and other suppliers respond to incentives, but there is little documented experience on creating incentives in public sector agencies for promoting and incorporating participation.

Third, is the use of participation justified—even if it is costly—simply because participation improves outcomes? Although a full estimate of the costs and benefits is beyond the scope of this article, it is a vital step in the research agenda.

Finally, can experiences with participation help to clarify the analysis of the deficiencies inherent in either a purely individualistic market or a purely statist, governmental approach to development? An analytic approach that incorporates participation might examine the various mechanisms whereby cooperative action by groups can overcome the inefficiency of individualistic solutions—for example, from free riding or strategic (mis)revelation of private information—while avoiding the limitations of centralized government. These informal methods of cooperation have been explored by several authors (Ostrom, Schroeder, and Wynne 1993; de Soto 1989; Wade 1988), but much remains to be learned.

Table A-1. *Coded Variables and Summary Statistics*

Variable	*Sample size*	*Mean*	*Standard deviation*
Performance indicators			
Overall project effectiveness	121	4.1	1.6
Percentage of water system in good condition	98	4.8	1.8
Objective value of benefits	120	4.2	1.3
Percentage of target population reached	118	4.9	1.1
Participation variable			
Overall participation	121	3.7	1.7
Fully exogenous performance determinants (Z)			
GNP per capita	114	519.8	389.3
Project complexity	121	3.3	1.2
Total cost (ln)	104	15.4	1.5
Adequacy of facilities	121	4.5	1.3
Difficulties in staff recruiting	92	3.8	1.7
Availability of parts	115	4.2	1.5
Target objectives	121	4.4	1.2
Other performance determinants (W)			
Appropriateness of technology	121	4.5	1.3
Support of government	118	4.6	1.1
Agency understanding	118	2.8	0.9
Conduciveness of political context	121	3.2	0.7
Conduciveness of economic context	121	3.2	0.7
Conduciveness of sociocultural context	121	3.5	0.7
Conduciveness of geological and environmental context	121	3.2	0.9
Average number of users	117	3.2	1.1
Competition from other sources	109	3.4	1.5
Skill of staff	111	4.6	1.2
Overall quality of management	120	4.2	1.3

Note: All qualitative variables are ranked on a scale of 1 to 7. GNP per capita and total cost are in dollars.

Source: Authors' calculations.

Table A-2. *Results of Bivariate, Limited Multivariate, and Full Multivariate Regressions of Project Performance, Using Various Models*

Variable	*Bivariate*	*Limited multivariate inclusive of all Z variables (no Ws)*	*Full multivariate inclusive of all Z and W variables*
Overall beneficiary participation	0.6	0.28	0.24
	(10.6)	(5.3)	(3.8)
Availability of parts		0.57	0.44
		(9.6)	(5.6)
Target objectives		0.22	0.04
		(2.9)	(0.4)
Adequacy of facilities		0.14	0.03
		(1.9)	(0.4)
GNP per capita		−0.0003	−0.00006
		(−1.5)	(−0.3)
Project complexity		−0.08	−0.07
		(−1.3)	(−0.9)
Difficulties in staff recruiting		−0.05	0.006
		(−1.1)	(0.1)
Total cost (ln)		0.04	0.08
		(0.8)	(1.3)
Appropriateness of technology			0.19
			(2.3)
Overall quality of management			0.21
			(1.6)
Support of government			0.10
			(1.1)
Average number of users			−0.08
			(−1.06)
Conduciveness of economic context			0.1
			(0.9)
Conduciveness of geological and environmental context			−0.1
			(−0.9)
Skill of staff			0.08
			(0.8)
Agency understanding			0.05
			(0.5)
Conduciveness of political context			0.03
			(0.24)
Competition from other sources			−0.01
			(−0.2)
Conduciveness of sociocultural context			0.02
			(0.1)
Sample size	121	77	68
R^2	0.49	0.86	0.89

Note: All qualitative variables are ranked on a scale of 1 to 7. GNP per capita and total cost are in dollars. The values are from OLS regressions in which the dependent variable is overall project effectiveness. *t*-statistics are in parentheses.

Source: Authors' calculations.

Table A-3. *First-Stage Regressions of Project Participation*

Instrument and variable	*All instruments*	*Extent to which participation was a goal*	*Percentage of investment costs paid by users*	*Net benefits of participation*	*Organization based on local collectives*	*Prior commitment of clients*
Instruments						
Extent to which participation was a goal	0.24	0.682				
	(2.89)	(12.9)				
Percentage of investment costs paid by users	0.07		0.505			
	(0.87)		(6.51)			
Net benefits of participation	0.248			0.658		
	(3.06)			(11.8)		
Organization based on local collectives	0.11				0.44	
	(1.47)				(5.83)	
Prior commitment of clients	0.044					0.49
	(0.452)					(4.03)
Exogenous variables in the limited multivariate regression						
Project complexity	−0.013	−0.034	0.158	0.014	0.068	0.151
	(0.191)	(0.431)	(1.40)	(0.176)	(0.73)	(1.19)
Total cost (ln)	−0.098	−0.019	−0.169	−0.029	−0.14	−0.286
	(1.73)	(0.329)	(1.93)	(0.474)	(1.99)	(2.86)
Adequacy of facilities	0.19	0.181	0.381	0.223	0.354	0.382
	(2.22)	(2.01)	(2.98)	(2.33)	(3.15)	(2.53)
Difficulties in staff recruiting	−0.016	0.041	0.004	−0.011	−0.052	0.009
	(0.323)	(0.75)	(0.051)	(0.192)	(0.783)	(0.108)
Target objectives	0.144	0.12	0.142	0.128	0.187	0.129
	(1.88)	(1.36)	(1.09)	(1.38)	(1.82)	(0.901)
Availability of parts	−0.048	−0.001	0.009	−0.086	0.003	−0.028
	(0.727)	(0.016)	(0.09)	(1.11)	(0.035)	(0.238)
GNP per capita	0.00051	0.00073	0.00034	0.00083	0.00043	0.00067
	(2.68)	(3.51)	(1.12)	(3.76)	(1.76)	(1.98)
Sample size	63	77	75	77	66	72
R^2	0.865	0.826	0.643	0.803	0.719	0.559

Note: All qualitative variables are ranked on a scale of 1 to 7. GNP per capita and total cost are in dollars. The values are from OLS regressions in which the dependent variable is overall project effectiveness. *t*-statistics are in parentheses.

Source: Authors' calculations.

REFERENCES

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Basu, Ritu, and Lant Pritchett. 1994. "The Determinants of the Magnitude and Effectiveness of Participation: Evidence from Rural Water Supply Projects." Background note for *World Development Report 1994*. World Bank, Washington, D.C.

Briscoe, John, and David de Ferranti. 1988. *Water for Rural Communities: Helping People Help Themselves*. Washington, D.C.: World Bank.

de Soto, Hernando. 1989. *The Other Path: The Invisible Revolution in the Third World*. New York: Harper and Row.

Esman, Milton J., and Norman T. Uphoff. 1984. *Local Organizations: Intermediaries in Rural Development*. Ithaca, N.Y.: Cornell University Press.

Finsterbusch, Kurt. 1990. "Studying Success Factors in Multiple Cases Using Low-Cost Methods." University of Maryland, Department of Sociology, College Park, Md. Processed.

Finsterbusch, Kurt, and Warren Van Wicklin III. 1987. "Contribution of Beneficiary Participation to Development Project Effectiveness." *Public Administration and Development* 7(January/March):1–23.

Freire, Paulo. 1973. *Education for Critical Consciousness*. New York: Seabury Press.

Greene, William. 1990. *Econometric Analysis*. New York: Macmillan.

Hammermesh, Daniel S., and Jeffrey E. Biddle. 1994. "Beauty and the Labor Market." *The American Economic Review* 84 (5, December): 1174–94.

Herren, Hans R. 1984. "Factors Associated with the Success of Participants in the National Future Farmers of America Livestock Judging Contest." *Journal of American Association of Teaching and Education in Agriculture* 25(1):1–19.

Korten, David C. 1980. "Community Organization and Rural Development: A Learning Process Approach." *Public Administration Review* 40(5, September-October):480–510.

Korten, Frances F., and Robert Siy, Jr., eds. 1988. *Transforming Bureaucracy: The Experience of the Philippine National Irrigation Administration*. West Hartford, Conn.: Kumarian Press.

Murphy, Kevin R., Robert A. Jako, and Rebecca Anhalt. 1993. "Nature and Consequences of Halo Error: A Critical Analysis." *Journal of Applied Psychology* 78(2):218–25.

Narayan, Deepa. 1995. "The Contribution of People's Participation: Evidence from 121 Rural Water Supply Projects." ESD Occasional Paper Series 1. World Bank, Environmentally Sustainable Development Department, Washington, D.C. Processed.

Ostrom, Elinor, Larry Schroeder, and Susan Wynne. 1993. *Institutional Incentives and Sustainable Development: Infrastructure Policies in Perspective*. Boulder, Colo.: Westview.

UNDP (United Nations Development Programme). 1993. *The Human Development Report 1993*. New York: Oxford University Press.

Wade, Robert. 1988. *Village Republics: Economic Conditions for Collective Action in South India*. Cambridge, U.K.: Cambridge University Press.

World Bank. 1991. *World Development Report 1991*. New York: Oxford University Press.

IRIS Reprint No.67

Does Participation Improve Performance? Establishing Causality with Subjective Data

Jonathan Isham, Deepa Narayan, and Lant Pritchett

IRIS Summary

An increasing number of development practitioners agree that participation of the intended beneficiaries improves project performance. Participatory development, championed since the early 1970s by mostly noneconomic social scientists and grassroots organizations (Freire 1973; Korten 1980), is increasingly advocated by the largest and most influential aid agencies (UNDP 1993; World Bank 1991). The existence of consensus (or advocacy), however, does not imply the existence of clear and convincing evidence that participation improves performance. Advocates of participation have most often relied on case studies to document the link between participation and performance (Briscoe and de Ferranti 1988; Korten and Siy 1988).

Skeptics have raised three objections to the statistical evidence of a causal impact of participation on project performance. First, participation and project ratings are not objective: subjective data are not appropriate for statistical analysis. Second, the subjectivity in the ranking of projects will lead to halo effects: if investigators believe participation is good, their subjective rankings will overstate the level of participation in highly successful projects and the level of success of highly participatory projects. Third, better project performance may cause increased beneficiary participation rather than vice versa: a mere statistical association is not evidence of a causal impact of participation on performance.

In this article we address and overcome each of the above objections. We start with a brief review of the construction of the data on project performance and beneficiary participation for 121 rural water supply projects and a presentation of the basic statistical results.

In answer to the first objection, we show that the subjective nature of the data does not preclude intersubjectively valid, cardinal measures of participation appropriate for statistical analysis. There is no necessary connection between objectivity and quantification with cardinal numbers.

We address the halo effect in coding performance and participation data from project documents in two ways. First, we show that the results are the same if the first coder's performance indicators are regressed on the second coder's participation scores (and vice versa); this test indicates that the coders' subjective ranking does not lead to a halo effect. The primary danger of subjective measurement, however, and one that we cannot address, is having the same individual assess both project success and participation. If that person has strong views about the relationship, these views may induce a bias in the project documents themselves, because the performance of participatory projects (and the participation in successful projects) may be exaggerated. We show, however, that the strength of the performance-participation relationship does not depend on the objectiveness of the success indicator.

The third and most difficult objection to answer is that the existence of an association does not imply causation. Although causality is nearly impossible to establish, we present three arguments in its

favor. First, estimation using instrumental variables allows the identification of the impact of exogenous changes in participation, and eliminates the effect on the participation estimates of reverse causation or simultaneity. Second, data on the timing of participation show that participation at early stages improves project performance at every stage, from implementation to maintenance. Third, case studies show that exogenous changes in participation in ongoing projects have strong impacts on performance.

We begin by showing the existence in project-level data of a strong association between project performance and beneficiary participation. We then address and answer the three econometric objections to acknowledging this association.

Four questions that are important for practice and policy are not explored here. First, does participation directly cause better project performance across all sectors? We cannot blindly apply the results in this study across all sectors, because these data are limited to rural water supply projects. The economic characteristics of rural water as a good would seem to promote the importance of direct beneficiary participation; these economic characteristics vary across goods provided by projects in other sectors.

Second, what policy instruments help to achieve more effective participation? Project beneficiaries, staff in project agencies, and other suppliers respond to incentives, but there is little documented experience on creating incentives in public sector agencies for promoting and incorporating participation.

Third, is the use of participation justified even if it is costly--simply because participation improves outcomes? Although a full estimate of the costs and benefits is beyond the scope of this article, it is a vital step in the research agenda.

Finally, can experiences with participation help to clarify the analysis of the deficiencies inherent in either a purely individualistic market or a purely statist, governmental approach to development? An analytic approach that incorporates participation might examine the various mechanisms whereby cooperative action by groups can overcome the inefficiency of individualistic solutions--for example, from free riding or strategic (mis)revelation of private information-while avoiding the limitations of centralized government. These informal methods of cooperation have been explored by several authors (Ostrom, Schroeder, and Wynne 1993; de Soto 1989; Wade 1988), but much remains to be learned.