

Evaluating the Social Effects of Performance-Based Environmental Programs

Cary Coglianese
University of Pennsylvania Law School

Jennifer Nash
Harvard Kennedy School

Jonathan Borck
Harvard Kennedy School

February 2008

The research underlying this report was conducted under EPA contract EP-W-05-047 with Industrial Economics, Inc. The analysis and views expressed herein are, of course, the authors' alone and not those of EPA, Industrial Economics, or the institutions with which the authors are affiliated.

Evaluating the Social Effects of Performance-Based Environmental Programs

Cary Coglianese
University of Pennsylvania Law School

Jennifer Nash
Harvard Kennedy School

Jonathan Borck
Harvard Kennedy School

Over the past decade, the U.S. EPA and some 22 states have established performance-based environmental programs (PBEPs) with the explicit goal of improving the environmental performance of private-sector facilities. As a condition for the recognition and rewards that come with membership in PBEPs, facilities must submit application materials to demonstrate that they meet specified environmental performance criteria. As part of their applications, facilities commit to improve their environmental performance in areas such as water and energy conservation and habitat protection. But even though PBEPs explicitly seek to improve the environmental performance of members and potential members, environmental performance is not the only goal. Many PBEPs seek to achieve a variety of social goals, such as enhancing trust in relationships between businesses, communities, and government agencies, or diffusing a more responsible ethos throughout corporate culture.

At two recent workshops about PBEPs organized by the U.S. EPA, the Environmental Council of the States (ECOS), and the Multi-State Working Group on Environmental Performance (MSWG), participants spoke about the importance of both of these social goals. Support for these social goals also showed up in recent interviews

conducted with EPA and state PBEP managers as part of a study of 18 PBEPs (Coglianese et al. 2008). Program managers with four PBEPs mentioned the goal of improving relationships without prompting, and managers at eight more agreed that it was a goal after being asked. When asked, managers of over 75% of the programs reported that their program goals included changing the culture of businesses and government agencies.

A report on the future of PBEPs, written by Allen White and Wei-Shiuen Ng of the Tellus Institute as part of the overall research initiative of which this paper is one part, calls on agencies to design future PBEPs with their social contributions in mind. The report argues that by emphasizing social goals, agencies will encourage facility managers to adopt “a more holistic, systemic view of how their facilities should be managed” (White and Ng 2008, 2). The report further recommends that agencies encourage facilities to adopt a consistent set of key performance indicators that would capture social as well as environmental performance.

According to membership surveys that EPA conducted in 2004 and 2006, Performance Track participants reported that social benefits provide much of the value they receive from participation. In response to the 2004 survey, members reported that their most important reason for joining and staying in the program was to enjoy a “collaborative/amicable relationship with EPA & States” (Abt Associates 2005, 4). The 2006 survey showed that most respondents believed that Performance Track participation had contributed to a “culture of continuous environmental improvement” and had improved their facility’s “relationship with EPA” (EPA 2006).

The primacy members place on social goals perhaps should not be surprising (cf. Moomaw 2001). Social goals are reflected in some of the requirements imposed on members of PBEPs. For example, the U.S. EPA's Performance Track and many state programs require that facilities have established a community outreach program as a condition for membership, and nearly all programs require facilities to inform the public on a regular basis about the environmental impacts of their operations. Nearly all agencies with established PBEPs offer opportunities for member facilities to engage with representatives of agencies and community organizations in meetings to share perspectives and best practices. If PBEPs bring facility managers, community residents, environmental advocacy organizations, and government environmental agencies into closer contact through meetings and other mechanisms for information sharing, the prevailing expectation is that traditional adversarial relationships will transform into more cooperative modes of interacting. The implications of such a shift could be substantial, as new perspectives and best practices begin to permeate facility walls and transform both business and government agencies (King 2006).

It is also conceivable, though, that PBEPs could negatively affect relationships and the regulatory culture. This could happen if PBEPs raise expectations for cooperation that are not met, or if repeated interactions provide greater opportunities for disagreement or miscommunication, with the possible unintended result of increasing levels of mistrust among stakeholders. Especially if agencies promise benefits to participants that they fail to deliver, or if facilities fail to live up to the commitments they make, relationships could be damaged rather than strengthened. It is also conceivable that government's stated commitment in most PBEPs not to subject members to routine

regulatory inspections could raise public suspicions and exacerbate public mistrust of both agencies and facilities. In 2005, when EPA moved to bolster benefits for Performance Track participants, the Natural Resources Defense Council submitted comments to EPA arguing that Performance Track amounted to little more than “a poker tournament,” trading away needed regulations and enforcement authority in the hope of encouraging facilities to cooperate (Walke 2005).

The possibility that PBEPs could fail in improving regulatory relationships and culture motivates the question we address in this short paper: How can agencies (or anyone) measure and evaluate PBEPs’ social contributions? Agencies have put a great deal of work into developing environmental performance metrics for PBEPs, collecting environmental performance data from member facilities, and compiling data into summary reports about these programs’ direct contributions to environmental protection (Coglianese et al. 2008). Much less attention has focused on establishing social impact measures and collecting social performance data (Coglianese et al. 2008). This paper therefore outlines what would be needed to demonstrate with confidence that PBEPs successfully achieve their social goals. While we focus on the social goals of improving relationships and regulatory culture, the issues we raise would apply equally to evaluating the contributions of these programs to achieving other social goals such as integrating environmental concerns into business decisionmaking and realizing consistent investment in activities that foster environmental protection.

Measurement Issues: Defining and Operationalizing Social Goals

An initial hurdle is to clarify what is meant by a “social goal” and develop appropriate performance measures. While environmental goals are themselves not always easy to define, for the most part they are the consequence of economic inputs and can be operationalized using outputs that can be isolated, measured, and tracked, such as emissions or energy usage. But what, exactly, does it mean to “improve relationships” among facilities, agencies, and surrounding communities? Or to “change culture” or “enhance trust”? Investigators cannot directly observe these phenomena: no “culture-o-meter” or “trust-o-meter” exists. In these cases, investigators must identify proxy variables that are correlated or associated with the underlying social variables of concern.

Proxy variables can fall into one of two categories: revealed proxies and expressed proxies. Revealed proxies are measures of actual observable behavior consistent with the underlying social variables of concern. Revealed proxies are closely related to the concept of “revealed preferences” in economics: they are measures of real-world behavior that reflect the preferences and attitudes of the actors. Examples of revealed proxies for “improving relationships” might be the number of complaints community residents make to or about an industrial facility or the number of lawsuits filed by groups against polluters.

In contrast, expressed proxies are measures of actors’ stated opinions. Expressed proxies are closely related to the concept of “stated preferences” in economics: they are measures of actors’ preferences determined by their words, not their actions. An example of a stated proxy for “improving relationships” might be the responses to a survey

question asking environmental or community organizations how much they trust industry or government regulators.

The problems with revealed proxies are that they are limited in availability and may be more closely related to effects other than the underlying variable of concern. The range of observable behavior is naturally limited. Businesses, regulators, and environmental and community groups engage in only a few observable actions plausibly related to the underlying social effects. And when they do, these actions are necessarily indirect measures of the underlying social variables of concern. Investigators must be aware of the possibility that the variation in a specific revealed proxy is related less to the underlying social effect than to some other motivation or influence. For example, the number of lawsuits filed by environmental groups against polluters could be explained more (or entirely) by changes in groups' finances than by changes in underlying levels of trust or improvements in relationships. Investigators do not need to abandon such revealed proxies: they can use statistical techniques to account for these alternative explanations. But the issue certainly complicates efforts to identify the effects of PBEPs.

Unlike revealed proxies, which are indirect and limited in availability, expressed proxies are readily obtainable and can be quite direct: investigators can ask directly about any underlying social variable of concern. But they simply cannot be sure that the responses they receive are accurate reflections of the "true" feelings of the respondents. Opinions obtained through surveys or interviews are prone to numerous biases that have been well documented in various literatures (Coglianese 2003; Tietenberg 2006). For example, respondents can be swayed by the range of options in survey questions or by the first option provided in an interview. Respondents might respond strategically to a

survey question to influence a perceived outcome. Respondents can also be influenced by the amount and nature of background information provided in a survey or interview or by external factors, such as the weather on the day they answer the survey or conduct the interview, that are entirely irrelevant to the questions being asked. Researchers are actively developing methods to attempt to elicit more truthful or less biased responses to survey or interview questions, and techniques have improved immensely from the earliest days of survey research. But implementing these improved methods to collect expressed proxies is both time consuming and expensive.

Another, more fundamental challenge with both revealed and expressed proxies is that, being they are proxies, the direction of any correlation between the proxy and the underlying variable of concern can be unclear: Is the proxy negatively or positively correlated with the underlying variable? For example, an “improved relationship” might mean one with less *conflict*, perhaps measured by the number of complaints community residents make to or about an industrial facility. However, increased *interaction* might also define improving relationships, in which case an increase in the number of complaints might signal first steps in developing relationships that will eventually overcome the problems or misperceptions that underlie the complaints. Similarly, greater *listening* could imply that researchers should count the number of hours facility managers spend in meetings with agencies and other stakeholders. But then again, lengthy meetings could also potentially indicate that facility and agency managers have reached an impasse and are no longer listening to each other. *Trust* might be shown through managers’ willingness to disclose large quantities of information about their environmental performance to community residents; on the other hand, being on the

receiving end of a large “data dump” could feel like bombardment and perhaps only reinforce feelings of mistrust. In each of these cases, investigators can observe plausible proxies for the underlying social variables of concern, but it may not always be clear whether changes in the level of the proxy indicate improvements or declines in the underlying social effect.

The desire to understand whether a PBEP leads to improvement gives rise to an added measurement challenge: namely, the proxies gathered to assess progress toward program goals should be collected not only from (or about) participants in the PBEP, but also from (or about) an appropriate sample of non-participants. Some revealed or expressed proxies may be readily obtainable from businesses, regulators, and environmental and community groups both within and outside a program. For example, investigators can presumably obtain a count of the number of lawsuits filed by community or environmental groups against polluters, whether or not those groups and polluters are involved in the PBEP. On the other hand, some proxies may be more difficult to obtain from (or about) non-participants. For example, facilities participating in a PBEP might readily respond to a program-sponsored survey asking them about trust in government regulators because of its obvious relevance, but non-participating facilities may be less likely to respond to a request to complete such a survey. Obtaining data from both types of facilities, however, is essential in order to draw proper inferences about the program’s impact, the subject of the next section.

Inference Issues: Determining the True Effect of a PBEP

Determining how well a PBEP achieves its goals involves more than just measuring progress toward the goals, even if appropriate proxy measures can be identified and collected. Investigators must use techniques to assess whether any progress they observe is actually the result of the program itself, or whether some other factor explains the results. They should be particularly aware of two confounding effects: the Hawthorne effect and omitted variables bias.

The Hawthorne effect draws its name from a study of worker productivity conducted in the early part of the last century at the Western Electric Company's Hawthorne facility in Illinois. The effect refers to the potential for study subjects to act or respond differently just because they know they are being studied. In the Hawthorne experiments, researchers varied working conditions (such as lighting, schedules, and so forth) in an effort to determine how these conditions affected worker productivity. They found that, over time, the productivity of the experimental group always increased—regardless of the changes made to workplace conditions (Roethlisberger & Dickson 1939). The Hawthorne effect reminds researchers that the very fact that members of a treatment group know they are part of an experiment and are being observed may influence how they perform or respond. The potential for the Hawthorne effect is undoubtedly inherent in any voluntary program like PBEPs. Researchers need to be mindful of the possibility that any effects that appear to have come from participating in *the* program under study may simply have come about because participating facilities knew they were in *a* program and were being observed.

The problem of “omitted variables bias” occurs when one or more factors left out of an analysis—usually because it is hard or impossible to observe them—have an important effect on the observed responses or outcomes. In such cases, the risk is that the effect of an omitted variable will be incorrectly ascribed to the variable measured and included in the analysis. For example, when investigators go to analyze the effects of participating in a PBEP on some proxy for trust of stakeholders, they may leave one or more important variables out of their analysis, perhaps inadvertently but more likely because it is difficult to gather data on all the variables that might influence the proxy. One such variable, for example, could be facilities’ pre-existing trust of other stakeholders. Not only might a facility’s pre-existing trust influence its trust as measured after it joined a PBEP—but that pre-existing propensity will also likely influence whether a facility joins a PBEP in the first place. If facilities’ pre-existing trust is not accounted for in a study of the social effects of PBEPs, any higher level of trust observed among participants will appear to have been caused by participation in the PBEP itself, when, in fact, some (if not all) of the difference was due to the pre-existing trust among participants. An analysis of trust will not reveal the true effect of participation in the program if the omitted variable is never taken into account.

One way to attack the omitted variables bias problem and isolate the true effect of a PBEP is to find a suitable proxy variable for the omitted variable. These proxies are slightly different from the proxies discussed in the section above. There, we discussed types of proxies for the underlying social effect of interest (or what we might call the outcome of concern). Here, we seek proxies for unobservable variables that influence the underlying outcome of concern. But the ideas—and the challenges—are the same

whether seeking proxies for the outcome of concern or for other variables that might affect the outcome of concern. For example, an expressed proxy variable for the omitted variable of facilities' "pre-existing level of trust" could take the form of responses to survey questions that had been administered before the PBEP was established. Under certain statistical conditions, this proxy variable can be used to control for the unobserved variable and help isolate the true effect of the PBEP.

Another way to tackle the omitted variables bias problem in a statistical setting is the so-called "instrumental variables" technique. Unlike a proxy variable, which is a variable correlated with the unobservable omitted variable, an instrumental variable is a variable correlated with participation in the PBEP. Under certain conditions, the instrumental variable technique can be used to isolate the true effect of the program. The method of instrumental variables is well known but challenging to implement, as it requires some known randomness in at least one factor affecting the voluntary decision to participate in a program like a PBEP (Bennear & Coglianese 2005).

A more straightforward way to tackle the omitted variables bias problem would be to use the so-called "differences-in-differences" method. Differences-in-differences is one of a class of statistical approaches that use data collected over multiple time periods. The differences-in-differences technique requires investigators to collect data on participants and non-participants in a program in two time periods: (1) before the program, and (2) after the program. The method assumes that participants would change over time the same way the non-participants did if they never joined the program. If so, then any *additional* change in the outcome variable or effect of concern among the participants can be inferred to have been due to the influence of the program itself. For

example, suppose that investigators observed that the reported level of trust among a sample of managers from participating facilities increased by a certain amount after the facilities joined a PBEP. Moreover, suppose the investigators observed that trust also increased among a sample of facilities that did not participate in the program over the same time period. The differences-in-differences technique would allow the research to conclude that the effect of the PBEP is the *additional* increase in trust among participants, not their total increase. The method provides not only a better estimate of the effect of the program, but also a more confident one: it helps to rule out alternative explanations, such as a pre-existing but unobservable propensity to trust.

Note that all the statistical techniques described above require gathering data about facilities (or other actors) that are not involved in the program. These non-participants serve as the “control group”—that is, they provide a basis for estimating what participants would have done in the absence of the program (the “counterfactual”) (Coglianese 2002). Without a properly chosen sample of non-participants against which to compare the behavior and evolution of participants, investigators cannot confidently estimate how much progress a PBEP is making toward its goals, social or otherwise.

The data needs are even more extensive to implement the powerful differences-in-differences technique. Investigators must collect four chunks of relevant data: data from (1) the participants before they joined the PBEP, (2) the participants after they joined the program, (3) a sample of non-participants before the participants joined the program, and (4) a sample of non-participants after the participants joined the program. Investigators who only collect data from participants in a PBEP after they have joined the program

have only collected one of the four chunks of data required to implement the differences-in-differences approach and thereby gain a best estimate of the true effect of the PBEP.

Of course, in mentioning these statistical techniques, we do not mean to imply that the social effects of PBEPs can only be studied through large samples using advanced quantitative analytic tools. Important insights can also be obtained through in-depth study of smaller numbers of facilities or programs. Yet even so, the challenges we have discussed about measurement and inference still arise. Fortunately, qualitative research can also be designed in ways that respond effectively to these challenges, such as with the use of carefully matched case studies (King, Keohane, & Verba 1994).

Linking Social Effects to Environmental Performance

Up to this point, we have been primarily concerned with social effects or outcomes themselves, as if these effects were intrinsically valuable. Perhaps they are, in that people may be happier to live and work in communities in which relationships among stakeholders are strong and cultures within businesses and government agencies are cooperative. However, there seems to be good reason to suppose that many policymakers and managers quite properly view social effects as simply a means to the larger end of improved environmental protection. After all, most agencies implementing PBEPs are “environmental protection” agencies, not “social capital building” agencies.

If social effects are important primarily because of their subsequent effects on environmental quality—for example, because firms that exhibit greater trust tend to take greater strides to improve their environmental performance—then identifying and

obtaining good measures of and inferences about social effects will not be enough. In addition to the measurement and inference challenges we have already discussed, researchers and policy analysts will also confront challenges in determining the association between social effects and environmental outcomes. Some of these challenges will be familiar. For example, to determine if any observed changes in social effects lead to changes in environmental outcomes, researchers will face measurement challenges, such as the need to identify and collect measures not only of social effects but also environmental outcomes. Unfortunately, not all environmental outcomes are regularly or reliably measured. Thus, investigators must use those measures that are available, such as the EPA's Toxics Release Inventory (TRI), as proxies for overall environmental performance. As with all proxies (as we discussed above), investigators must be aware that the proxies may not be highly correlated with the underlying outcome of concern, in this case overall environmental performance. Reported TRI figures, for example, may provide at best only a partial indicator of facilities' aggregate levels of pollution and their overall environmental performance (especially when that performance is understood to include energy and water use, among other things).

Familiar issues of inference also complicate the linking of social outcomes to overall environmental performance. As is true for any research issue, correlation is not the same as causation. Just because investigators observe that facilities with higher social variables also have superior environmental performance does not mean that the increase in social variables led to the superior environmental performance. For one thing, superior environmental performance might be what leads to increases in social variables, not the other way around. Cleaner firms may prompt, and generate for themselves, greater trust.

Furthermore, an unobserved third variable—the familiar “omitted variable” described above—may be responsible for both observed effects in the social and environmental variables. In other words, there may be something else—perhaps effective managerial leadership—that both promotes greater trust *and* improved environmental performance. Researchers and policy decision makers must be particularly careful to consider, and try to rule out, alternative explanations and causal pathways before crediting improvements in multi-stakeholder relationships, increases in trust, or changes in organizational culture with any observed improvements in environmental performance.

Conclusions: A Path Forward

Evaluating the effects of *any* public program—not just PBEPs—requires attention to the kind of issues we have discussed in this paper. These issues also arise no matter what the goals of the program may be. For this reason, no one interested in PBEPs should think that the challenges in evaluating these programs will necessarily be any easier when the goals are defined in social rather than environmental terms. Any well-executed and meaningful evaluation of PBEPs—whether for their impact on social or environmental goals—will need to attend to these concerns. Unfortunately, it will never suffice simply to poll program members to see if they are satisfied with the program or if they think it is having social or environmental effects (Coglianese 2003). These effects need to be demonstrated through careful empirical research that attends the issues we have outlined in this paper.

Lest we be misunderstood: Even though such research can be a daunting task, it is not impossible. Our discussion of evaluation issues in this paper suggests ways to design evaluations that will yield convincing results. For example, investigators should first take care to identify plausible proxies for the underlying social effects of concern. Identifying and collecting multiple proxies for a single social effect or outcome can increase the level of confidence that the proxies are measuring the right effect, especially when the direction of correlation between a single proxy variable and the underlying social effect or outcome is unclear.

If investigators are primarily interested in social effects as a means to improvements in environmental performance, they should take care to link the social effect to environmental performance explicitly. To improve the confidence in making inferences about the effects of PBEPs, investigators should be sure to collect data both from participants in the program and an appropriate sample of non-participants. Since some of the most powerful inference techniques require data from before the participants joined the program, investigators may wish to focus evaluations on industrial sectors or groups of facilities that have not yet joined a particular program and follow them as they join. Taking these evaluation issues into account before establishing new PBEPs could provide opportunities for collecting pre-program data that can be used to compare to post-program outcomes or responses. Overall, investigators must be aware of and transparent about alternative explanations for any correlations they observe. Only by addressing the issues we have outlined in this paper can researchers rule out alternative explanations for their results and thereby increase confidence in what they can conclude about PBEPs' success in achieving their goals, social or otherwise.

References

- Abt Associates, Inc. 2005. Results of 2004 Performance Track Customer Satisfaction Survey. Available at: http://www.epa.gov/performance-track/members/news/mar05/survey_report.pdf, accessed February 1, 2008.
- Benear, Lori and Cary Coglianese. 2005. "Measuring Progress: Program Evaluation of Environmental Policies," *Environment* 47(2): 22-40.
- Coglianese, Cary, Jennifer Nash, and Jonathan Borck. 2008. State and Federal Performance-Based Environmental Programs: Assessing Goals, Activities, Communication, and Data Collection. Research conducted under EPA contract EP-W-05-047.
- Coglianese, Cary. 2003. "Is Satisfaction Success? Evaluating Public Participation in Regulatory Policy Making," in Rosemary O'Leary and Lisa Bingham, eds., *The Promise and Performance of Environmental Conflict Resolution*. Washington, DC: Resources for the Future Press, 69-86.
- Coglianese, Cary. 2002. "Empirical Analysis and Administrative Law," *University of Illinois Law Journal* 2002: 1111-1137.
- King, Andrew. 2006. "The Role of Management Systems in Stakeholder Partnerships." In *Leveraging the Private Sector: Management-Based Strategies for Improving Environmental Performance*, Cary Coglianese and Jennifer Nash, eds. Washington, DC: Resources for the Future Press, 228-245.
- King, Gary, Robert O. Keohane, and Sydney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
- Moomaw, William R. 2001. "Expanding the Concept of Environmental Management Systems to Meet Multiple Social Goals." In *Regulating from the Inside: Can Environmental Management Systems Achieve Policy Goals?*, Cary Coglianese and Jennifer Nash, eds. Washington, DC: Resources for the Future Press, 126-145.
- Roethlisberger, F.J. and William J. Dickson. 1939. *Management and the Worker*. Cambridge, MA: Harvard University Press.
- Tietenberg, Tom. 2006. *Environmental and Natural Resource Economics*. Seventh Edition. Boston: Pearson Addison Wesley.
- U.S. Environmental Protection Agency. 2006. 2006 Performance Track Survey: Supplemental Questions Preliminary Results. Washington, DC: U.S. EPA.

Walke, John. 2005. Letter to Docket ID OA-2005-0003, November 3, 2005.

White, Allen, and Wei-Shiuen Ng. 2008. Visioning the Future of Performance-Based Environmental Programs. Research conducted under EPA contract EP-W-05-047.