Evaluating the Performance of Collaborative Environmental Governance

Craig W. Thomas
Associate Professor
Daniel J. Evans School of Public Affairs
University of Washington
Seattle, WA 98195-3055
thomasc@washington.edu

Prepared for presentation at the Consortium on Collaborative Governance Mini-Conference, Santa Monica, April 10-12, 2008.
Abstract

Collaboration has increasingly supplemented and other forms of environmental governance, such as centralized planning and command-and-control regulation. Hence, practitioners and academics routinely debate whether collaboration improves the environment over alternative governance systems. But the debate is largely rhetorical and theoretical, because there is little empirical evidence to suggest whether collaboration has a positive or negative impact on the environment. This paper reviews the current state of research on collaborative governance, and suggests ways to design research studies that test the links between collaborative processes and environmental outcomes. The paper also argues that collaborative governance should be held to environmental performance standards, just like other governance systems.
Introduction

Is collaborative governance good for the environment? Practitioners and academics routinely ask this question, because collaboration has increasingly supplemented and supplanted other forms of governance, such as centralized planning and command-and-control regulation. Collaboration is one of several forms of new governance elbowing its way into the environmental policymaking process. Hence, practitioners and academics routinely debate whether collaboration improves the environment over alternative governance systems. But the debate is largely rhetorical and theoretical, because there is little empirical evidence to suggest whether collaboration has a positive or negative impact on the environment.

Proponents argue that collaborative governance is well suited for addressing certain types of environmental problems, such as non-point-source pollution (John 1994), ecosystem management (Cortner and Moote 1999; Grumbine 1994), and biodiversity preservation (Thomas 2003). But that does not mean that collaboration actually produces better environmental performance than other governance systems. The crux of the argument lies in implementation. Proponents argue that collaboration reduces conflict and litigation, leads to shared ownership and authority, increases trust among stakeholders, increases community capacity to address problems, and leads to better management of natural resources (U.S. GAO, 2008). According to this line of reasoning, collaboration produces more durable solutions because participants buy into the policies they recommend, adopt, or implement. Critics counter that collaborative policies tend to be least-common-denominator solutions, particularly if consensus is the primary mechanism for reaching decisions (Coglianese 1999). Thus, one of the key trade-offs in choosing collaborative governance is whether the benefits of durable solutions (during implementation) outweigh lower common-denominator solutions (during policy adoption). Unfortunately, we lack evidence to demonstrate the magnitude – and direction – of this trade-off.

Most research on collaborative environmental governance focuses on processes (e.g., consensus, public participation, and mediation). Some research also addresses outputs (e.g., plans and projects) and social outcomes (e.g., trust and social capital). Very little research addresses environmental outcomes. Thus, we know much about how collaborative processes vary (Koontz, et al. 2004; Sabatier, et al. 2005). Emerging research also suggests how collaborative outputs vary (Wilkinson 2007; U.S. IECR 2007; U.S. GAO 2008), and how collaborative outputs differ from non-collaborative outputs (Leach 2007). Research also indicates that collaboration on environmental problems tends to increase social capital among participants (Sabatier, et al. 2005; U.S. IECR 2007), though we still know little about the affect of collaboration on non-participants. One recent study, for example, suggests that increased social capital among collaborative participants may come at the expense of reduced social capital in other venues (Lubbell 2007). In sum, we know much about what collaboration is and what collaborative partners do, but we know very little about the impact of collaboration on environmental conditions.
The dearth of research on collaborative environmental outcomes reflects, in part, the life cycle of collaborative partnerships. Most of the collaborative environmental partnerships that have been studied emerged after the 1980s; hence, there has been insufficient time for many of these partnerships to generate outputs, let alone monitor the effect of these outputs on the environment. We know much about the state of the environment; but relatively little about how specific policies, programs, and governance systems affect environmental conditions. Research on environmental outcomes has lagged well behind research in other fields, such as health outcomes (Heinrich and Fournier 2004), education outcomes, fire prevention outcomes (Donahue 2004) and human services outcomes (Page 2004), just to name a few.

Environmental outcomes are difficult to assess. Endangered species are a case in point. We know much about how the Endangered Species Act (ESA) is implemented, but relatively little about how ESA implementation affects populations of endangered species. In part, this is because numerous confounding factors affect species’ populations other than the ESA, such as climate change, invasive species, and disease. The problem becomes even trickier when we ask whether it matters if the ESA is implemented through collaborative planning and management or some other governance system (such as command-and-control regulation, centralized planning, market-based mechanisms, or a combination of these governance systems). Thus, not only is it difficult simply to show how ESA implementation affects environmental outcomes, it is even more difficult to analyze whether collaboration on ESA implementation (through, for example, habitat conservation plans) has a positive effect on environmental outcomes. An exception to the problem of attribution in environmental policy is point-source pollution (from pipes and stacks), because the cause-effect relationship between regulatory enforcement, point-source emissions, and ambient air quality is relatively straightforward.

Ideally, if we could design research protocols to evaluate the environmental performance of collaboration, then we could build this knowledge into a performance management system that holds collaborative partners accountable. Collaborative accountability is already occurring for human services outcomes in several states (Page 2004). Recent research on other environmental governance systems, such as voluntary environmental management systems, suggests that expectations of environmental performance tend to exceed actual performance (Coglianese and Nash 2006; Khanna 2007). It is time to empirically assess the environmental performance of collaborative governance as well.

This paper reviews the state of knowledge on the environmental performance of collaborative governance, and proposes methods to fill in knowledge gaps regarding environmental outcomes. While some published research has claimed to demonstrate causal links between collaborative processes and environmental outcomes, this paper will demonstrate that these publications overstretch the data. The goal of this paper is to put research on environmental outcomes on a solid footing, and to suggest how research on environmental performance can be integrated into performance management systems.
Evaluating Environmental Performance

Program evaluation has been around for decades, but the language and methods of program evaluation have made few inroads into the field of environmental policy (Bennear and Coglianese 2005), and even fewer inroads in the subfield of collaborative environmental governance. In this paper, “collaborative environmental governance” refers to any local, state, or federal effort to solve an environmental problem within partnerships among public, private, and nonprofit organizations. Collaborative partnerships may be influenced by other governance systems, such as hierarchy or contracting, but the primary feature of collaborative governance is these partnerships. Collaborative governance can be evaluated in the same way that policies and programs are evaluated, by using a logic model that carefully distinguishes collaborative processes from the outputs and outcomes of those processes.

Logic models specify the causal impacts one expects will occur as a result of a policymaking process. Logic models typically contain the following components (Hatry 2006), which can arranged sequentially in a causal chain, as depicted in Figure 1.

**Inputs**: Resources used in a process (e.g., the technical, financial, and human resources contributed by partners in a collaborative process).

**Processes**: Workload or activities (e.g., collaborative meetings).

**Intermediate outputs**: Early products and services delivered from a process (e.g., collaborative agreements or plans).

**End outputs**: Subsequent products and services delivered from a process (e.g., projects in a plan, and monitoring and enforcement of those projects).

**Intermediate outcomes**: Conditions outside a process that precede the desired end result (e.g., changed human behavior near wetlands due to plans and permits).

**End outcomes**: The end result sought (e.g., aquatic conditions in a wetland).
Logic models can be much more complex than depicted in Figure 1. For example, outputs can be decomposed into longer chains of outputs (e.g., draft plans, final plans, implementation agreements, permits, and implemented projects). Additional causal arrows can be added to account for various feedback loops. Rather than attempting to model the causal sequence in great complexity, we should focus on the key components associated with successful performance and then develop metrics for each component (Hatry 2006). In this paper, successful performance refers to positive environmental outcomes.

Modeling the logical structure of a collaborative effort provides two primary methodological benefits. First, logic models specify what to measure. Rather than simply gathering available data, logic modeling requires us to think carefully about what should be measured, and the validity of existing measures. The purpose of logic modeling is to be logical rather than expedient – that is, to develop measures based upon a logically derived model, not the expedient collection of existing data. A second methodological benefit is that logic models specify the expected causal relationships between the characteristics of an organized effort and the effects of that effort. For example, if a collaborative process contains characteristics A, B, C, then the logic model tells us to expect outputs L, M, N and outcomes X, Y, Z. These hypothesized relationships can be deduced from prior knowledge (academic theory or past practice) or from professional intuition (informed hunches about what is likely to occur).

In terms of performance assessment and management, logic modeling before implementation compels us to state precisely what “success” would mean in practice. Post-hoc rationalizations are less likely to occur if an explicit logic model exists prior to implementation. If a collaborative partnership defines “success” in terms of outcomes
prior to implementation, then subsequent backsliding to measures of outputs or process in performance appraisals will be readily transparent.

More generally, logic models can be embedded in a performance management system that links evaluation with accountability (Hatry 2006; Imperial 2005; Behn 2004). This is rarely done in practice for environmental issues. Notable exceptions will be covered later in the paper. Some local, state, and federal agencies are experimenting with environmental performance management systems, but these systems vary in the quality and extent of their logic modeling, and none yet demonstrate the benefits of collaborative governance over other governance systems.

Evaluating the Performance of Collaborative Environmental Governance

Existing research on collaborative environmental governance has focused on process, outputs, and social outcomes. Very little is known about environmental outcomes. While some publications claim to measure environmental outcomes, they typically use outputs as proxies for outcomes. Hence, the link between collaborative processes and environmental outcomes remains elusive.

Evaluations of Collaborative Inputs and Processes

Much is known about the inputs that support (or hinder) collaborative processes on environmental issues. Existing research on inputs covers a wide range of topics, including the role of government institutions and actors (Koontz et al. 2004), scientific knowledge (Thomas 2003), and local stakeholders (Weber 2003; Sabatier, et al. 2005). This literature also covers multiple scales, including large-scale regional collaborative efforts (Layzer forthcoming; Heikkila and Gerlak 2005; Thomas 2003) and small-scale watershed partnerships (Sabatier, et al. 2005; Weber 2003).

The existing literature linking inputs and processes is largely explanatory, rather than evaluative. It explains the antecedents to sustainable collaborative efforts, emphasizing the role of beliefs (Sabatier, et al. 2005), trust (Leach and Sabatier 2005), and institutions (Koontz et al. 2004; Thomas 2003; Ostrom 1990). Some new research on collaborative environmental processes is self-consciously evaluative. The Institute for Environmental Conflict Resolution is completing a study of 52 cases in which a neutral third party was used in reaching an agreement. In addition to evaluating collaborative processes, this study also evaluates collaborative outputs and social outcomes (U.S. IECR 2007).

Evaluations of Collaborative Outputs

Recent studies have measured collaborative outputs, and evaluated collaborative processes in terms of outputs. The U.S. GAO (2008), for example, recently analyzed the outputs of collaborative resource-management processes, but the findings cannot be easily generalized. The GAO study sampled mostly successful cases, thus providing little variation between successful and unsuccessful partnerships. The outputs coded for each
partnership also differed, making comparison difficult for assessing causal relationships among even the successful efforts. Thus, the GAO study indicates the types of outputs associated with successful collaborative efforts, but does not tell us what unsuccessful collaborative efforts produce, or how collaborative outputs differ from non-collaborative outputs.

In a more systematic study, Wilkinson (2007) coded the contents of 12 watershed plans and 15 salmon recovery plans in Washington State. His findings demonstrated that partnership plans varied widely in their recommendations. Some of the plans offered clear recommendations, while others offered only general recommendations that did not articulate clear actions for moving forward. The findings indicate that collaborative partnerships produce plans that are highly varied in the robustness of their recommendations, even when the partnerships operate under very similar guidance from public agencies. These recommendations must still be implemented before they have an effect on outcomes. Nonetheless, this study provides systematic data on intermediate outputs that can be used in subsequent evaluations of end outputs and outcomes.

A study nearing completion by Layzer (forthcoming) provides seven case studies of large-scale ecosystem conservation and restoration efforts, which vary in the amount of collaboration. She finds that collaboration hindered effective environmental management. More collaboration produced less protective outputs, which is consistent with the critique that collaboration generates lower common-denominator solutions.

In addition to these studies, which self-consciously evaluate outputs, other studies implicitly measure outputs but mislabel them as outcomes. One accordingly must read these studies carefully to interpret the findings. Koontz, et al. (2004), for example, use the term “environmental management outcomes” to encompass both outputs and outcomes, when in fact most of the measures in the book are outputs. Weber (2003), by comparison, uses the term “outcome” as a synonym for “choices.” These choices can be collaborative agreements or other decisions, and hence are outputs in the language of logic modeling.

Other publications use measures of outputs as intentional (rather than accidental) proxies for outcomes. Daley (2007) uses output measures (Superfund site remedies) as a proxy for environmental outcomes in her analysis of the impact of citizen participation. Though the title of her article refers to environmental outcomes as the dependent variable, she uses output measures as proxies because “there are no comparable and agreed upon measures” of environmental quality or change at Superfund sites (Daley 2007:353). Meyer and Konisky (2007) similarly use output measures (plans and permits) as proxies for environmental outcomes in their study of the impacts of local environmental institutions. Like Daley, they note in the text that these are output measures, but nonetheless claim that their dependent variables are outcomes because outputs are good proxies for outcomes (Meyer and Konisky 2007:486-488).

The confusion could be easily remedied if authors would simply claim that outputs are the dependent variables in these studies, not outcomes. It would also help to note whether the outputs are intermediate or end outputs. In studies that focus on
collaborative choices (Weber 2003), these outputs occur early in a chain of outputs. Studies that focus on plans and permits (Meyer and Konisky 2007) observe outputs further in a chain of outputs. Permits could be the end output in a logic model if the next step is behavioral changes by permit holders; but it would be more appropriate to add monitoring and enforcement as end outputs prior to behavioral changes. Behavioral changes are outside the collaborative process, and are thus intermediate outcomes. The end outcomes would be changes in environmental conditions. Plans and permits are thus weak proxies for end outcomes, because many other factors affect environmental conditions, including enforcement. Moreover, outputs may have unintended consequences that harm the environment.

In sum, an expanding literature on collaborative outputs is emerging. Some of these studies carefully distinguish outputs from outcomes. Other studies conflate outputs with outcomes or use outputs as proxies for outcomes. The language of program evaluation arrived late to the field of environmental policy (Bennear and Coglianese 2005). Thus, authors have been less familiar with the careful distinctions that should be made between outputs and outcomes in logic modeling, and have been overly eager to claim that their data indicate the effect of collaborative processes on environmental outcomes.

Evaluations of Collaborative Outcomes

Collaborative outcomes can be divided into two types: social and environmental (Koontz and Thomas 2006). Social outcomes are the primary performance metric in fields such as human service delivery (Page 2004). Environmental partnerships also produce social outcomes, such as trust and social capital (Leach and Sabatier 2005; Lubell 2005; Lubell 2007). Environmental outcomes include changes in environmental conditions, such as water quality, ambient levels of pollution, and biodiversity. This paper focuses on environmental outcomes, because the literature on social outcomes is already well developed. We can also learn much about social outcomes from research on collaborative governance in other fields.

Environmental outcomes remain the elusive Holy Grail of environmental performance. While some publications have claimed to evaluate collaborative environmental outcomes, they do so by using outputs as proxies, as noted above, which means they are really evaluating outputs. Conflating outputs with outcomes may lead to erroneous conclusions about the actual environmental impacts of collaboration (or any other governance system). This is not simply an academic debate about terminology – there are serious policy implications. For example, a recent literature review in *BioScience* (Svancara, et al 2005) found that policy-driven approaches focused on outputs (e.g., collaborative agreements driven by benchmark targets of acreage to be protected in a country) led to significantly less protection than evidence-based targets focused on environmental outcomes (e.g., the acreage a species needs to thrive or survive). But another way, we need evidence of outcomes to drive choices of outputs, not evidence of outputs as proxies for outcomes. If land is being set aside for species protection, then the
appropriate measures should be biological outcomes, not land set aside for environmental protection.

Leach and Sabatier (2005; 2003) have conducted the only known research so far on collaborative environmental outcomes. They gathered extensive empirical evidence of participant perceptions of environmental improvement, not actual environmental conditions. These authors are well aware of the methodological shortcomings in using this measure of environmental outcomes. First, the measure is of perceptions, not actual environmental conditions. Second, participant perceptions are colored by a “halo effect” that leads participants to perceive better environmental conditions if they perceive their collaborative effort in a positive light. Thus, we still need objective measures of environmental outcomes, and methods for testing the causal links between collaboration and environmental outcomes.

Research Designs for Studying the Environmental Outcomes of Collaborative Governance

The primary challenges for evaluating environmental outcomes involve finding appropriate data, allowing for long time horizons, and developing research designs sufficient for establishing attribution (Koontz and Thomas 2006; Sabatier, et al. 2005, pp. 277-280). Researchers need data sets that provide valid measures of the environmental outcomes, as predicted by a logic model. It should be time-series data that includes years preceding the collaborative intervention to establish a benchmark for improvement that can be attributed to the collaborative effort. Long time horizons are therefore needed to allow collaborative outputs to have effects on environmental conditions. The data should also allow for comparisons between collaborative and non-collaborative processes. Remote sensing data is particularly promising in these regard (Schweik and Thomas 2002). One then needs research designs that can distinguish the impact of collaborative partnerships from other causal factors affecting environmental conditions, as noted below.

Collaborative partnerships typically do not gather such data themselves (Koontz and Thomas 2006, p. 114); and it would be “a herculean task” to expect collaborative partnerships, particularly small partnerships, to monitor and measure their own environmental performance (Sabatier, et al. 2005, p. 278). Moreover, it is better to have third-party evaluators monitor performance to maintain neutrality. Therefore, the onus should be on academics and professional evaluators, rather than collaborative partnerships themselves. The remainder of this section discusses several research designs for testing the links between collaborative processes and environmental outcomes.

Experimental Designs

Randomized trials have become the gold standard in the fields of public health, education, and social policy; but not in environmental policy. This discrepancy can be readily seen in a recent issue of The Annals of the American Academy of Political and Social Science (Boruch 2005), which includes numerous articles reporting randomized
trials in the fields of public health, education, criminology, and social programs – but none from environmental policy. Some environmental professionals argue that it is too difficult to conduct classic experimental designs with regard to environmental policy, and that environmental programs should not be held to this standard (Dawes 2008). Regardless of the technical challenges, there is little reason to conduct them from a purely pragmatic perspective due to the long time horizon required for environmental change following randomized trials of collaborative and non-collaborative processes. For policy and management purposes, it is better to have some knowledge sooner rather than better knowledge later.

Quasi-Experimental Designs

Quasi-experiments are the most promising research design for attributing environmental outcomes to collaborative processes (or any other governance system). Quasi-experiments lack one or more of the components of classic experimental designs, usually means random assignment (Shadish, Cook and Campbell 2002). The absence of random assignment raises several threats to validity, but these threats can be assessed using a variety of methodological techniques, and are potentially offset by practicality. The great benefit of quasi-experimental designs for evaluating collaborative environmental outcomes is that one can choose cases in which the long time horizons necessary for environmental change have already occurred. One need only find data to measure environmental outcomes.

Remote sensing is a useful tool for measuring environmental change, because monitoring data extends back to the 1970s. This long time horizon extends well before many collaborative processes began, and the data is collected at regular intervals for the same area (or footprint). Remote sensing can thus provide reliable and valid measures of environmental change that can be attributed to institutionalized processes (Schweik and Thomas 2002). Moreover, remote sensing can measure a wide variety of environmental conditions, including water quality, not just land cover change.

Testing whether collaboration improves environmental outcomes over alternative governance systems requires finding matched case comparisons. Mill’s method of difference (also known as the most-similar method) is well suited for this purpose. In this method, one chooses cases that are alike in as many respects as possible, except for the variable of primary interest. Thus, one could vary the governance system while holding other factors constant. While the cases are not randomly assigned, one still controls for other factors by holding them relatively constant in the matched cases. Mill’s method of difference could thus be used in quasi-experiments (with remote sensing or other measures of environmental outcomes) to compare the impact of collaboration with other governance systems. For example, one could compare collaborative with non-collaborative habitat conservation plans (HCPs), using the ECOS data set of HCPs maintained by the U.S. Fish and Wildlife Service (ecos.fws.gov) to find matched cases. Or one could similarly search for matched cases of community fire plans, national fish habitat partnerships, BLM resource advisory committees, or other processes that vary in
collaboration (e.g., the type of decision-making process, the number of collaborative partners, or other indicators of collaborative inputs).

Case-Study Designs

Case-study designs differ from quasi-experiments in that they rely heavily on process tracing to understand causal mechanisms (Gerring 2007; George and Bennett 2005). Thus, they are particularly effective for tracing how the dominoes fall in long causal chains in logic models. While quasi-experiments show whether particular environmental outcomes are associated with particular governance processes, they cannot show why these associations occur. Quasi-experiments show causal effects, not causal mechanisms.

In a case study, one analyzes each step in a causal sequence, using a variety of techniques to demonstrate the causal mechanisms that operate between variables. Techniques for process tracing include the congruence method, controlled comparisons, and counterfactual reasoning (George and Bennett 2005). These techniques are well suited for demonstrating the causal mechanisms leading from inputs to processes to outputs to outcomes. Case-study designs are also well suited for demonstrating specific trade-offs noted earlier in this paper, such as the hypothesized trade-off between lower common-denominator solutions (during policy adoption) and more durable solutions (during implementation). In a case study, we can trace the implementation process to see whether and why implementation is more durable in a collaborative process.

Statistical Designs

Statistical designs can be used by themselves or in combination with the other designs noted above. The key feature of statistical designs is examining correlations between variables. Statistical controls substitute for experimental controls (whether randomized controls in classic experiments or matched-case controls in quasi-experiments and comparative case-study designs). Statistical methods can also be used to provide supporting evidence of causal mechanisms. Statistical methods provide only supporting evidence because they test correlations, not causation.

Using statistical methods, one could mine, for example, mine the databases cited in the Program Assessment Rating Tools (PARTs) submitted by federal agencies to OMB (Radin 2006; Hatry, et al. 2005). PARTs can be easily found on OMB’s ExpectMore.gov website (http://www.whitehouse.gov/omb/expectmore). Each PART contains (or should contain) measures of program outputs and outcomes. One could then test statistical correlations between measures of processes, outputs, and outcomes. Unfortunately, PARTs vary greatly in quality, so one would have to be very careful in selecting PARTs that report valid measures of outputs and outcomes. Some PARTs, for example, confuse output and outcome measures. Thus, statistical methods should be used in tandem with case study methods to ensure that the indicators and statistical conclusions are valid.
A Path Forward

This paper has explored the current state of research on collaborative governance, and suggested ways to design research studies that explore the links between collaborative processes and environmental outcomes. Existing research has provided us with much knowledge about the links between inputs and collaborative processes. Recent research is also building knowledge about the links between collaborative processes and outputs. But there is very little research thus far on the links between outputs and environmental outcomes.

I am currently working with Tom Koontz (Ohio State University) and a research assistant on several projects to push this research agenda forward. These projects include case studies of public-sector efforts (at the local and federal levels) to demonstrate the links between program processes and environmental outcomes. We are also coding PARTs as a prelude to finding matched case comparisons for statistical analysis. We also continue our work with Clare Ryan (University of Washington) coding the outputs of watershed partnerships, as a prelude to linking outputs with outcomes. Our plan is then to use remote sensing data and ecological data as measures of environmental outcomes.

In the long run, collaborative governance should be held to environmental performance standards, in the same way that other governance systems and programs are held to performance standards. To now, most of the debate about costs and benefits of collaborative governance (in terms of environmental performance) have been rhetorical and theoretical. Emerging research supports the hypothesis that collaborative governance produces less environmental protection than other governance systems, but insufficient evidence exists to demonstrate rather this translates into reduced environmental conditions. It might be that collaborative governance generates more durable implementation, as proponents argue, which may offset lower protective actions. We will not know the answer for some time, but we need the evidence to hold collaborative partners accountable for their actions.

An emerging literature calls for performance-based accountability in regulatory governance (May 2007), voluntary compliance (Coglianese and Nash 2006; Morgenstern and Pizer 2007), and conflict resolution (U.S. IECR 2007; Brogden 2003; Rowe 2003). It is time to follow this trend by assessing the environmental performance of collaborative governance (Imperial 2005; Koontz and Thomas 2004; Weber 2003).
References


