



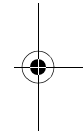
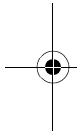
9

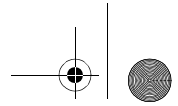
Evaluation of Environmental Dispute Resolution Programs

ANDY ROWE

Evaluation is a vehicle for acquiring quality information about the performance of programs and activities. Systematic evaluation is an integral element of reflective practice and an invaluable contribution to improved practice and theory of dispute resolution. As a unified field of practice, program evaluation, like dispute resolution, is relatively young, and we are still developing the knowledge and practice necessary for high levels of performance. Evaluation offers a mechanism for systematically gaining valid and reliable information about dispute resolution—information that can successfully undergo critical scrutiny from internal and external stakeholders.

Because some of the claims and statements I will make in this chapter could be surprising, I believe that it is important for readers to know my background in both evaluation and working with dispute resolution programs. I have a long and successful track record in evaluation and limited recent experience developing evaluation systems for two different programmatic settings—the U.S. Institute for Environmental Conflict Resolution (USIECR) and the Oregon Public Policy Program, of which the Oregon Dispute Resolution Commission is one of three state agency partners. My evaluation experience spans almost 20 years of practice in diverse public program areas, including environmental and resource programs, all areas of health and human services, arts and culture programs, and international development. Over this period I have employed most of the many approaches evaluators have in their toolkits. I was president of the Canadian Evaluation Society (CES) for two years and have been privileged to serve at national levels in both the CES and the American Evaluation Association (AEA). One thing





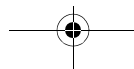
that has struck me over the relatively brief period in which I have been involved with dispute resolution programs is the gap between professional program evaluation and evaluation of dispute resolution. From anecdotal and personal observation, it seems that many who evaluate dispute resolution programs come from some association with the practice area itself and have limited evaluation experience outside these programs. As indicators of this, when I asked attendees at the invigorating and successful conference associated with this volume how many were members of the American Evaluation Association or had attended AEA meetings, only a few indicated that they fitted the bill. Likewise, an important state dispute resolution program held a set of invited meetings on evaluating dispute resolution programs the same week as the AEA annual meetings. Both examples demonstrate the gap between professional evaluation and the evaluation of dispute resolution programs. Both practices could benefit from closer collaboration among practitioners.

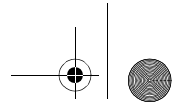
Because of this observed gap, my initial focus in this chapter is to communicate what evaluation can offer to dispute resolution, and then I assess the options for evaluating environmental dispute resolution programs. The chapter concludes with a summary of the key opportunities and challenges. Before embarking on that agenda, it is important to set evaluation in its contemporary context.

Contemporary Context for Program Evaluation

The expectations and requirements of public and nonprofit funders changed fundamentally in the 1980s and 1990s. This period saw the enactment of federal and state legislation requiring that programs focus on and achieve success in attaining their mandated results. The movement began in the United States, Australia, and New Zealand, but by the end of the period could be seen in all industrialized and many developing countries. Foundations and other nonprofit funders also enacted similar requirements. In the United States, the Government Performance and Results Act (GPRA) has changed forever the expectations and requirements of accountability for federal agencies and programs. Most states have also enacted accountability and performance requirements for state agencies, many before GPRA—for example, the Oregon legislature requirement of performance measures from all state agencies.

This move to a results focus and accountability has raised the profile and utility of evaluation for state, federal, and community agencies. It has also provided a useful structure around which evaluators can organize their work. In 1997, John A. Koskinen of the U.S. Office of Management and Budget described GPRA this way:





At its simplest, the Government Performance and Results Act (GPRA) can be reduced to a single question: what are we getting for the money we are spending? To make GPRA more directly relevant for the thousands of federal officials who manage programs and activities across the government, GPRA expands this one question into three: what is your program or organization trying to achieve? How will its effectiveness be determined? How is it actually doing? One measure of GPRA's success will be when any federal manager anywhere can respond knowledgeably to all three questions.

The evaluation systems I developed for the Oregon Public Policy Program and the U.S. Institute for Environmental Conflict Resolution directly address these three questions.

Evaluation and Dispute Resolution Programs

Many readers of this volume will not be familiar with the categories and terms used in program evaluation. I will outline a few terms that are basic to evaluation of dispute resolution programs.

Evaluators have traditionally grouped evaluations as *summative* or *formative*.¹ Summative evaluations are about judging the merit or worth of a program or service; the purpose of formative evaluation is to support program improvement. Evaluation is reaching a level of maturity where the distinction between formative and summative evaluation is of less value. At one time, summative evaluation was taken to refer to assessment of outcomes, whereas formative evaluation assessed processes. This was never really the case and today is clearly not correct; the difference has more to do with the types of decisions made using the evaluation information than with evaluation methods or focus.

The agenda for evaluation of dispute resolution programs is both summative and formative. It includes summative considerations of whether to fund these programs or use more traditional mechanisms of addressing disputes (including doing nothing), and it is formative because it provides information identifying means of improving the effectiveness of particular dispute resolution programs, or indeed, the practice of dispute resolution. Most existing evaluations of dispute resolution programs appear to have a formative intent; however, legislatures and foundations are now asking questions and considering decisions that are best addressed with summative evaluations. Ideally, programs should first have the benefit of formative evaluation to aid them in achieving higher levels of performance and as a resource for any summative evaluations that might be requested.



This points to a second consideration regarding the purpose of evaluation. Evaluation seeks to acquire information that is good enough to support current and likely future decisions about the program. This is the guiding principle for the quality of evaluation information and can be contrasted to the standards for applied research that would engage peer review, publication, and hopefully, use. Referring to the program decisions that are likely to be made using evaluation information provides a useful general guide to the appropriate design of an evaluation.

Before initiating an evaluation, an *evaluability assessment*² should always be undertaken. Not all programs should be evaluated; for example, programs in their early stages are unlikely to be performing near their potential, and it would be appropriate to use evaluation approaches for immediate program improvement purposes, but inappropriate to use evaluation to judge the merit or worth of the program at this stage. In assessing a program's evaluability, evaluators should take into account

- whether the program has established clear and observable outcomes,
- if adequate information either exists or can be obtained with a reasonable level of effort,
- if there is good evidence that key stakeholders will use the evaluation information, and
- if sufficient key stakeholders can be engaged in the evaluation process.

Dispute resolution programs appear no more or less likely to be evaluable than programs in general. The three programs included in the collaborative evaluation initiative reported on elsewhere in this volume³ easily passed an evaluability assessment.

As evaluation practice and theory developed, the concept of a *program theory* has gained importance and today is central to many approaches to evaluation. Program theory is the theoretical and practice knowledge underpinning a program. When designing evaluations, evaluators increasingly want to establish the conceptual strength of the program's claims at an early stage. Surprisingly, many programs are launched and continue without ever articulating the underlying program theory, with the near-certain effect of reduced effectiveness and efficiency. Unfortunately many evaluators still undertake an evaluation without first clarifying the program theory, thereby enhancing the likelihood that the evaluation will expend resources assessing program activities that could have been made more efficient if the underlying program theory had been clarified. Program theory is also indispensable in allowing evaluators to offer evaluation judgements about the wider contributions of the programs. For example, many international development programs are funded from sources that have the goal of poverty reduction, but the funded programs themselves do not contribute directly to poverty reduction. Improved environmental infrastructure (such as water, sewage,



and solid waste disposal) in slums does little to reduce poverty directly, so we use program theory to provide the link. So long as the environmental infrastructure program successfully achieves the program theory, evaluators can make the statement that the program “is likely to” contribute to poverty reduction, based on good contemporary and relevant knowledge from practice and theory.

The strength of the program’s claims can first be assessed via the program theory, and this also provides a valuable guide to the quality of the information the evaluation requires: what is good enough? Lee Sechrest, a professor of psychology at the University of Arizona—Tucson, uses an effective and simple illustration. For some diseases, inoculation has been established as the optimal treatment approach. This is based on extensive epidemiological research, itself testing hypotheses resting on well-established physiological, pharmacological, and other disciplinary knowledge. From this we know that as long as we monitor inoculations and the conditions under which they are provided, we can be reasonably sure that the desired outcome will be achieved. Consider an alternative treatment—the “laying on of hands.” The theoretical and empirical foundations of this treatment are not as well established and do not engender the same level of confidence about program success that we associate with inoculation. Thus, the required quality of the evidence to evaluate an inoculation program is modest. For this we track outputs: who was inoculated where and with what? The claims made by “laying on of hands” can be regarded as quite exceptional, requiring exceptional evidence to evaluate program effects. This would likely require complex experimental designs sustained over a long period and conducted in a variety of settings for a range of ailments.

When I have stated that dispute resolution has a reasonably robust program theory, the response has ranged from outright disagreement to grudging, “Okay, but it is not really a program theory.” Perhaps with the advantages of a fresh perspective I have been able to see the practice of dispute resolution more clearly—or perhaps my experience has been too limited, and I have not yet encountered the programs and dispute resolution practices that would put the lie to my claim. Illustrated below are the primary outcomes that dispute resolution seeks to achieve.⁴ These were developed with Susan Brody and Mike Niemeyer of the Oregon Public Policy Program (Oregon PPP) and are highly congruent with the outcome structure developed with the U.S. Institute for Environmental Conflict Resolution (USIECR). The differences between the Oregon PPP and the USIECR structures lie in program emphasis, delivery, and accountability, not in the underlying program theory. There is also a tight match between these two outcome structures and the program conception developed by David Fairman of the Consensus Building Institute with the Massachusetts Office of Dispute Resolution. All three were developed independently, as part of a collaborative effort provid-



ing credence to the claim that the dispute resolution practice pursues a fairly homogeneous set of outcomes.

When one considers the strength of the underlying knowledge from literature on organizations, psychology, and negotiation, it seems reasonable to state that there is a fairly homogeneous program theory. This does not by any means imply that practice is homogeneous; that is clearly not the case. The program theory states the outcomes that diverse practice approaches seek to achieve.

Figure 9-1 groups process and agreement outcomes separately and distinguishes two further subsets of process outcomes: alternative dispute resolution is used appropriately, and party interactions are constructive. Both process and agreement outcomes are recognized and valued in this program theory, and there is an assumption that successful process outcomes are necessary to have successful agreement outcomes. The top gain from the use of dispute resolution processes is that parties and agencies can redirect resources that were allocated to the conflict issues to more productive uses.

The logic model is one means of describing a program and articulating a programmatic theory. There are many variants of a logic model. The one pictured here is partial but typical, but there are many different ways that evaluators describe outcomes, impacts, and results. Contemporary evaluation places considerable emphasis on the right-hand side of this figure. The key terms in a logic model are activity, output, outcome, and some term that indicates higher-level outcomes that are beyond the ability of the program to affect alone. Figure 9-2, which also includes impacts, uses the following terms: *activities* are what the program does (facilitate resolution of disputes), *outputs* are the products of the activities (number of disputes facilitated), *outcomes* are what has changed as a result (parties collaborate, no unresolved issues), *results* are the biggest outcome(s) for which the agency is accountable, and *impacts* are how society benefits from the program (fewer resources expended on disputes).

Logic models have become much in vogue—public and foundation funders now frequently require fundees to develop a logic model. This is a welcome step, but like most developments that come into vogue, form often becomes more important than content.

Evaluation Methods and Environmental Dispute Resolution Programs

To achieve the goal of judging a program's merit or worth, evaluation must be able to make valid and reliable statements about the program's effects compared to what would have occurred without benefit of the program. I refer to this as the *incrementality* of the program. Evaluation uses the entire

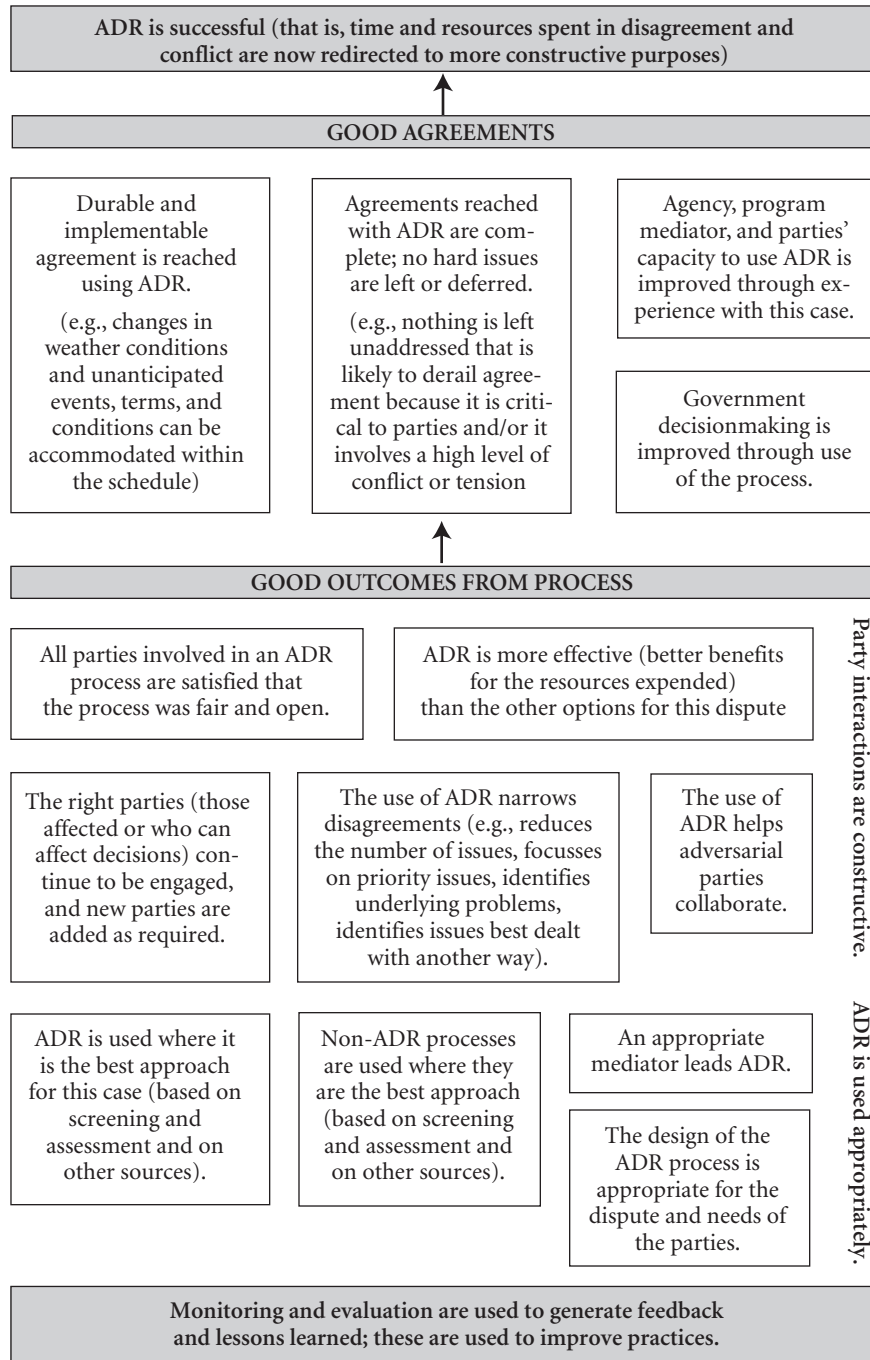


FIGURE 9-1. Oregon Public Policy Program Case Management Options

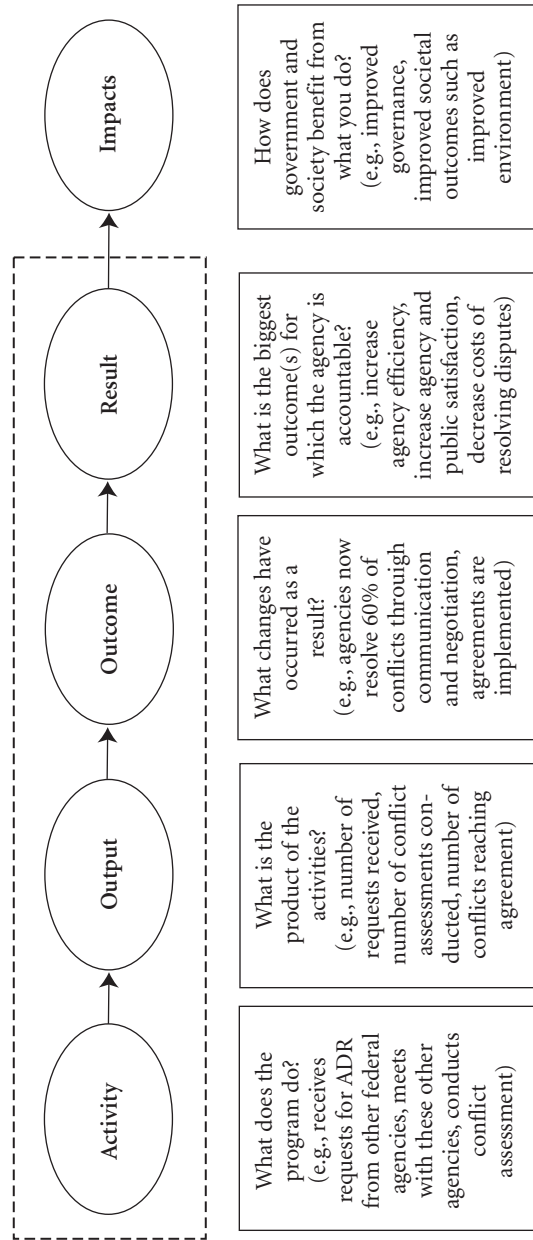
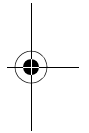


FIGURE 9-2. Core Definitions in a Logic Model

Note: The program is accountable for what lies inside the dotted-line box. In achieving these, the program contributes to impacts, but it is not accountable alone for achievement of impacts.





range of social science research methods, and incrementality can be assessed using qualitative or quantitative approaches,⁵ or a combination of both. Dispute resolution is a diverse practice with differences in approaches and in programmatic settings, requiring different evaluation approaches. The Program Evaluation Standards,⁶ developed by the Joint Committee on Standards for Educational Evaluation, direct evaluators to ensure that their work is appropriate, useful, of good quality, and affordable.

I have heard a number of stories about external evaluators requiring programs to randomly assign disputes to a dispute resolution treatment group or to a comparison group where dispute resolution services are not provided and the parties use other available options. These experimental and quasiexperimental designs can be appropriate in some settings, but should not be employed where they are not the optimum approach. Experimental and quasiexperimental designs rely on statistical techniques that require members of the program population to have an equal or known likelihood of being in the treatment group. The treatment and nontreatment groups are then both monitored for program effects, and the presumably positive differences are attributed to the program. The general requirement for this approach is a relatively large number of fairly homogeneous disputes and either homogeneous program treatments, or, at least, known differences in program treatment along with knowledge of the disputes' distinguishing characteristics. This approach is the most direct means of identifying incrementality, it has a high level of validity, and it is a well-established approach in social and natural research. I have used these designs frequently and to good advantage. If such methods were appropriate, I would usually employ a quasiexperimental design with a group of comparison projects, rather than an experimental design with random assignment of cases to the treatment or nontreatment groups. For programs such as dispute resolution, random assignment has no identifiable benefit over comparison groups and runs the risk of reducing program effectiveness because cases are often assigned to dispute resolution on the basis of some explicit or implicit screening to ensure a match between the case and the treatment applied.

Environmental disputes are typically quite heterogeneous because of important characteristics of the disputes and the parties. Moreover, the context within which dispute resolution is considered can significantly alter disputes; for example, a court's mandate to use dispute resolution likely alters how parties approach the resolution process. To date, the only settings I have encountered where the requirements of a quasiexperimental design might be met are in family courts, human resource complaints, or similar large-scale programs with fairly homogeneous cases, where sufficient and appropriate program data exists for all disputes in the study population. Given the complexity and heterogeneity of environmental disputes, it is extremely unlikely that they would ever meet the requirements for this approach. This means



that another evaluation design is required to evaluate environmental dispute resolution cases. Where quasiexperimental design is not suitable, many evaluators look for comparators. For example, we might seek to match participants in an employment program to nonparticipants with similar characteristics. For environmental disputes, we would seek to find similar disputes that were addressed by a reasonably good alternative, such as a judicial settlement conference, rulemaking, arbitration, litigation, or doing nothing. Some studies ask parties and neutrals to compare their dispute to other alternatives and use this information to make evaluative judgements about the incrementality of dispute resolution. This information is not sufficiently valid or reliable to support these claims—parties rarely have sufficient experience and knowledge of alternatives, and even when they do, the comparisons they make are unlikely to be to similar disputes. Without comparators, it is difficult to credibly state the incremental contribution of dispute resolution. This is a significant challenge for the evaluation of environmental dispute resolution processes, which I will return to in the next section.

Regardless of the means by which evaluators establish the incrementality of the program, many research methods can be used. Good case studies are appropriate for some settings, while others are better served through systematic surveys, anthropological methods, or collaborative and empowerment approaches. No approach has inherent superiority—good evaluators will use the approach that provides credible information that is good enough for the decisions that are likely to be made.

Evaluators must be careful to ensure that their evaluation efforts do not get in the way of the program. This is particularly important in dispute resolution programs, where success is highly dependent on the relationship between the neutral⁷ and the parties and the relationships among the parties themselves. The presence of evaluation could potentially disturb this relationship. It is my current view that evaluation should “stay out of the room” unless it is certain that it will not affect the dispute resolution process. Collaborative and empowerment evaluation approaches offer the best options for getting evaluation “inside the room” (see opportunities below).

Many existing evaluations of dispute resolution programs are best described as case studies from a legal or process perspective, using interviews, document reviews, and court information, or are limited reviews of party satisfaction and program implementation. There have also been efforts to use retrospective interviews with parties and others to obtain information about the dispute and the resolution process. My limited review of existing evaluations was neither systematic nor thorough, so it is with caution that I offer the observations that although these studies have been useful and important, the incremental contribution of dispute resolution has not been established, and the quality is often not good enough for the usual types of decisions that would use evaluation information.



Challenges and Opportunities

Systematic evaluation of dispute resolution programs offers valuable inputs for improving the performance of these programs and a means of responding to the increasing accountability requirements of public and foundation funders. Many of the benefits from evaluation will be realized early in the evaluation process. Both the Oregon PPP and the USIECR report benefit from their early efforts to clarify program outcomes and accountabilities. In addition, analysis of the early data from the evaluation has already identified some potential areas for improving overall performance of the programs and neutrals. Prospects are strong that the evaluation systems will also identify aspects of dispute resolution practice that can be improved. However, there are some important challenges in evaluating dispute resolution programs and environmental dispute resolution programs.

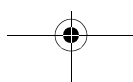
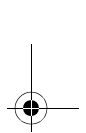
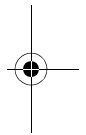
Challenges

The main challenge in evaluating environmental dispute resolution programs—because of the difficulty in identifying and obtaining information from comparators—is demonstrating the incremental benefit of the program. The second challenge applies to dispute resolution programs in general: obtaining valid and reliable information on the effectiveness of the program.

Incrementality. Assessing a program's incremental contribution is at the core of evaluation and presents a serious challenge for programs like environmental dispute resolution that feature extreme heterogeneity and complexity. The key is to assess the incremental benefit of the dispute resolution approach compared to other good alternatives. For this we require good information about what would have happened if one of those alternative approaches had been used instead of dispute resolution.

I have briefly described some of the approaches evaluation uses to do this. However, individual environmental disputes appear to be sufficiently unique that comparators are difficult to find. Whereas there are many water disputes, and indeed many water disputes involving similar classes of parties, they can still differ critically in terms of the parties' characteristics, the approach, the way that the neutrals approach the parties, the process actually used, the contextual setting for the dispute, and many other ways. Still, if we had good information on a large number of water disputes with similar parties, the importance of these differences would diminish, but we do not.

However, all is not lost. Early information from the Oregon PPP suggests that parties consider nonlitigative options such as legislation and administrative rulemaking as preferred alternatives to dispute resolution. Furthermore, in some jurisdictions (such as Oregon), attorneys from the Department of





Justice estimate the likely settlement and the internal costs of the state litigating larger cases and compare those estimates to calculated alternative dispute resolution costs. Thus, there are places to start looking for comparator alternatives to environmental dispute resolution. The requirement is for typical types of disputes and disputes addressed through the most likely means, and it is likely that for environmental disputes, the number of cells in this matrix of exemplars would not be that great. However, without applied research to obtain this information, evaluation is constrained in estimating the incrementality of dispute resolution approaches, something with which most external stakeholders such as legislators and funders are very concerned.

Measuring the Effectiveness of Dispute Resolution Programs. Dispute resolution has sometimes been promoted as “faster, better, and cheaper.” Although I am told that these claims are no longer made as often or as enthusiastically as they once were, they still have considerable intuitive appeal to legislators and other funders. Indeed it is reasonable to expect any program, when employed appropriately, to generate a combination of greater benefits (better) and greater efficiency (faster and cheaper) compared to the alternatives. This is what we mean by program effectiveness—more benefits per unit of input. Programs can increase their effectiveness by improving the amount or quality of the benefits they achieve for participants, reducing the amount of input required by improving efficiency in the provision of the programs, or a combination of the two. Thus the statement “faster, better, and cheaper” could be restated as, “Dispute resolution, when used appropriately, is more effective than the alternatives.” This is a reasonable claim that bears assessment.

It is difficult to measure the effectiveness of dispute resolution. On the cost side, there are serious measurement challenges, some general to estimating the types of costs involved in disputes and some that arise from the duration of disputes. It is also difficult to measure the benefits of dispute resolution.

Conceptualizing and Measuring Costs. When we measure costs of a given dispute addressed with dispute resolution processes, we want to do it in a manner that we can replicate: similar disputes should enable comparison of the different program alternatives’ effectiveness. This is essential for assessing the effectiveness of dispute resolution initiatives. On a larger scale, this will allow us to assemble a body of knowledge about ways to reduce the costs to parties participating in dispute resolution processes, as well as increasing accountability. There are two deep challenges in this. First, disputes, and often dispute resolution processes, occur over a long period of time. Disputes can exist for years, even decades or longer, before engaging dispute resolution or another alternative. Moreover, it can take many years for all the benefits and costs of a dispute to occur. To properly measure the costs of a dispute, we



want to compare the cost before the use of dispute resolution processes to the cost after all of the benefits occurred (regardless of whether an agreement was reached and implemented). This difference, plus the direct costs of resolving the dispute, constitutes the net change in costs attributable to the program intervention. We assume that the net change should be negative—the costs associated with the dispute should go down, and the fall in costs should be greater than the expense of the dispute resolution process itself.

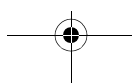
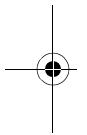
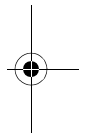
The longevity of disputes is only one of the challenges. Another challenge arises when we try to conceptualize the costs of disputes. Figure 9-3 provides a partial enumeration of the types of costs that could be associated with the dispute. It is challenging to conceptualize many of these sufficiently to design a standardized information-gathering approach.

There is also a serious measurement problem with implicit costs, such as staff and volunteer time, and with estimating the efficiency and effectiveness of an organization. My own previous timekeeping experiments with consulting staff, for example, indicated that they systematically underestimate billable time by 15% when recording time at the end of the day in 15-minute intervals. I found that their records were even more inaccurate on a weekly basis.

These are important challenges for evaluation, but they are not insurmountable. Remember that evaluation information only needs to be good enough for the decisions that are likely to be made. Evaluation needs are often satisfied with information from applied research. For example, when I was evaluating a geological mapping program as part of a mineral development initiative, with the benefit of known ratios applicable to the area that had been mapped, I was able to roughly estimate the likely future value of economic benefits. The key was the empirically established ratio of 1 mine per 25 exploration sites. Geologists were able to forecast the high and low numbers for exploration sites from the actual results of the mapping, and I used economic input–output tables to estimate the economic benefits of the mines and hence of the mapping. The quality of the forecast was good enough for us to estimate the economic benefits of the geological mapping, but clearly would not be sufficient for publication in an economics journal. Similarly, dispute resolution needs a few known and credible coefficients from which implicit costs can be reliably estimated.

Like the geological mapping example, we can use the extensive applied research on environmental impacts and other applications to generate good enough estimates of the costs (and benefits) to the environment that would arise from most agreements. Likewise, we can estimate other economic effects, such as employment and earnings, that are likely to arise from resolution of a dispute.

The problem is that we do not have the benefit of applied research that would provide us with the necessary coefficients to estimate with sufficient



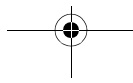


Upstream Costs	Transaction Costs	Downstream Costs
<p>Examples of Explicit Costs</p> <p>Fees of attorneys and other professionals</p> <p>Charges attributable to the dispute</p>	<p>Examples of Explicit Costs</p> <p>Fees of attorneys and other professionals</p> <p>Charges attributable to the dispute resolution process</p>	<p>Examples of Explicit Costs</p> <p>Fees of attorneys and other professionals</p> <p>Charges attributable to implementation of an agreement, continuation of a more limited dispute, or the entire original dispute.</p>
<p>Examples of Implicit Costs</p> <p>Time of staff and volunteers</p> <p>Reduced organizational efficiency and effectiveness because of engagement in the dispute</p> <p>Gains such as earnings that do not occur because of the dispute</p>	<p>Examples of Implicit Costs</p> <p>Time of staff and volunteers</p> <p>Outcomes that do not occur because the organization is involved in resolving the dispute</p>	<p>Examples of Implicit Costs</p> <p>Time of staff and volunteers</p> <p>Changes in organizational efficiency and effectiveness</p> <p>Any changes in gains (positive or negative)</p> <p>Costs of environmental damage, including costs to parties other than the disputants</p>

FIGURE 9-3. Classification of Costs of a Dispute

standardization the costs of a dispute to individual parties. This leaves evaluators of environmental dispute resolution programs in a quandary. We can expect that respondents will systematically underestimate current indirect costs if asked and that this will be more serious when we seek information over longer time periods. This creates a bias against dispute resolution being measured as efficient.

Estimating Benefits. The benefits of dispute resolution are described in Figure 9-1 for the Oregon PPP and can be grouped into dispute-specific benefits and party benefits. Dispute-specific benefits are improved collaboration among disputing parties; narrowing of the number of disputed issues; parties' satisfaction with the process, which is more effective than the alternatives; improved capacity of parties to engage in collaborative processes; com-





plete, durable, and implementable agreements reached; and improvement of government decisionmaking. According to the Oregon PPP program theory, all of these outcomes are desirable, and both process outcomes and agreement outcomes are valued. The Oregon PPP has three additional outcomes that the use of alternative dispute resolution must provide in addition to those identified in the program theory described in Figure 9-1: decreased costs of resolving disputes, increased efficiency of agencies, and increased agency and party satisfaction. These three outcomes result from the enabling legislation and are much on legislators' minds. Together these outcomes articulate the intended benefits from the Oregon PPP.

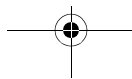
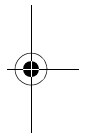
Both the Oregon PPP and the USIECR distinguish between process and agreement outcomes, and the program theories state that the process outcomes are necessary for achievement of agreement outcomes. The early evidence from the Oregon PPP and the USIECR supports this claim.

We can group the benefits from the dispute resolution process as improved capacity of parties to work together with other parties now and in the future and as complete, durable, and implementable agreements that will result in improved efficiency and effectiveness of the parties. Capacity and agreement benefits can be assessed with information from the parties and neutral(s); however, assessing the effectiveness benefits has the same challenges described above. It is likely that the most important benefit will prove to be improved organizational effectiveness, enabled by a reduction in the number and intensity of disputes requiring senior and middle management attention.

Summary of Challenges to Evaluating Dispute Resolution Programs.

Every program should be able to substantiate that it is the most effective option when used appropriately, and dispute resolution is not different. However, there are some serious challenges to obtaining the necessary cost and benefit information for disputes. These challenges arise both for general reasons and for reasons that are specific to dispute resolution. The general problems arise from the difficulties in getting valid and reliable information on the types of implicit costs and benefits that are involved in dispute resolution. The problems specific to dispute resolution arise from the extremely long time over which costs and benefit information must be obtained. Resolution of this challenge will come from applied research that generates knowledge of the costs and benefits for different classes of parties for the main types of environmental disputes.

The main methodological challenge is assessing the incremental contribution of dispute resolution compared to other good alternatives for addressing a dispute. This challenge is a serious constraint on our ability to respond adequately to summative evaluation questions about the value of dispute resolution. When combined with the additional challenges of measuring effective-





ness, evaluators of environmental dispute resolution programs face a potentially daunting task. However, it is a challenge that must be addressed because legislators and other funders have an obligation to fund programs that are the most effective response to a problem. To make these funding decisions in the context of accountability for results, funders require information about the effectiveness of dispute resolution programs compared to reasonable alternatives.

Fortunately, the gap in our current knowledge can be addressed with modest investments in applied research to generate dispute resolution exemplars and comparators for a limited number of types of disputes and classes of parties. In the meantime, evaluation is already showing itself to be a means of generating valuable information that can be used to improve environmental dispute resolution programs.

Opportunities: Collaborative, Participatory, and Empowerment Evaluation

Evaluation information is useful to the parties and neutrals facilitating the dispute resolution process, as well as to program officials, and collaborative and empowerment evaluation approaches are worthwhile experiments to improve the utility of evaluation to key stakeholders, neutrals, and parties. These approaches engage the program participants (parties and neutrals) in all aspects of the evaluation as primary stakeholders. Use of these approaches would require capacity building with the neutrals and parties and offers the promise of collaboratively set and implemented monitoring of the process and outcomes of dispute resolution. The information that would be obtained would give parties resources to improve collaborative dispute resolution processes in which they participated and would provide ongoing monitoring of any agreements, as well as information they could use when considering whether to engage in dispute resolution or other options for future disputes. These evaluation approaches need to be embedded in the program processes themselves. They offer the potential to enhance the benefits of evaluation to parties while obtaining better information about the dispute resolution process and its effectiveness, because collaborative and empowerment approaches would bring evaluation “inside the room” and inside the dispute resolution process itself.

Conclusion

This chapter has approached evaluation of environmental dispute resolution programs from the vantage of the professional evaluator. Good evaluation of environmental dispute resolution processes would clarify dispute resolution



programs and practices, as well as the mechanisms to gain information about programs and practices. The mechanisms can be used to identify and plan improvements in the theory and practice of environmental dispute resolution and to address the accountability requirements of legislators and funders. However, evaluation of environmental dispute resolution programs is constrained by the twin challenges of establishing the incremental contribution of dispute resolution compared to alternative processes and measuring the effectiveness of dispute resolution. Both of these challenges can be addressed with modest investments in applied research. Evaluation, enhanced with this knowledge, can fulfill the potential of not only aiding improvements in environmental dispute resolution practice and theory, but also aiding program managers, dispute resolution professionals, legislators, and other funders to use dispute resolution where it is the best option and ensure that key stakeholders benefit from the process.

Notes

I am grateful to everyone at the U.S. Institute for Environmental Conflict Resolution, the Oregon Public Policy Program, collaborators in the Policy Consensus Initiative, and in particular to Dale Keyes, Kirk Emerson, Mike Niemeyer, and Susan Brody for their patience and sharing their tremendous insights and knowledge. But of course, despite their contributions, I still make mistakes, and they are entirely of my own doing.

1. See Michael Scriven, *Evaluation Thesaurus*, 4th ed. (Newbury Park, CA: Sage Publications, 1991).

2. See Joseph S. Wholey, "Assessing the Feasibility and Likely Usefulness of Evaluation," in Joseph S. Wholey, Harry P. Hatry, and Kathryn E. Newcomer, eds., *Handbook of Practical Program Evaluation* (San Francisco: Jossey-Bass, 1990), pp. 15–39.

3. See Chapter 10 in this volume.

4. Note that further outcomes are nested under the principal outcomes and that there has also been an effort to define and clarify terms, for example, see the websites of USIECR, www.ecr.gov (accessed January 17, 2003) or the Indiana Conflict Resolution Institute, www.spea.indiana.edu/icri (accessed January 17, 2003).

5. The American Evaluation Association has established topical interest groups (TIGs) around evaluation methods and topics. The list of TIGs testifies to the diversity of evaluation practice and concerns and can be found at the AEA website, www.eval.org (accessed January 17, 2003).

6. Joint Committee on Standards for Educational Evaluation, *The Program Evaluation Standards*, 2nd ed. (Thousand Oaks, CA: Sage Publications, 1994).

7. I use neutral as a generic term to refer to the individual(s) who facilitate the dispute-resolution process. Through experience, I am mindful of the minefield on which I tread by doing so. See Scriven, *Evaluation Thesaurus*.

