

Margoluis, R., Stem, C., Salafsky, N., & Brown, M. (2009). Design alternatives for evaluating the impact of conservation projects. In M. Birnbaum & P. Mickwitz (Eds.), *Environmental program and policy evaluation: Addressing methodological challenges. New Directions for Evaluation*, 122, 85–96.

# 8

## Design Alternatives for Evaluating the Impact of Conservation Projects

Richard Margoluis, Caroline Stem, Nick Salafsky,  
Marcia Brown

### Abstract

Historically, examples of project evaluation in conservation were rare. In recent years, however, conservation professionals have begun to recognize the importance of evaluation both for accountability and for improving project interventions. Even with this growing interest in evaluation, the conservation community has paid little attention to evaluation design. Recent literature includes some discussion of design, but it has focused primarily on experimental and quasi-experimental design and the use of counterfactuals. Real-life conservation projects, however, operate in complex and dynamic contexts and under conditions of limited resources, which limit the feasibility of counterfactual or experimental designs. There is, in fact, a range of design options to evaluate conservation interventions. The conservation community must educate itself about these options, and how different designs produce different results. This chapter discusses evaluation design alternatives in light of the unique challenges that influence evaluation design selection in conservation. © Wiley Periodicals, Inc.



The conservation profession has matured to the point at which it must demonstrate positive results and wise use of conservation dollars. No longer is the adage “trust us; we are doing good work” sufficient to satisfy supporters and society at large. Given both the general calls for accountability and the stakes involved in conserving resources for future generations, we can ill afford to be inefficient in selecting interventions; we must learn and document what works and does not work in order to improve conservation decisions.

Despite all the thinking and experience in program evaluation over the past decades, the conservation community has not successfully applied this knowledge to conservation evaluation. Many researchers and practitioners recognize the utility of systematic evaluation, yet they rarely apply it. In fact, there is scant evidence of the use of the range of evaluation approaches regularly applied in fields such as public health, community development, and education (Stem, Margoluis, Salafsky, & Brown, 2005).

Although there has been some recent discussion of evaluation design in the conservation literature, much of it has centered on the role of experimental and quasi-experimental design and use of counterfactuals (Ferraro & Pattanayak, 2006; and Ferraro, this issue). But typically, real-life conservation projects operate in complex and dynamic contexts and under conditions of limited resources. In these real-world situations, it is often not feasible to use counterfactual or experimental designs (Donaldson & Christie, 2005; Mark & Henry, 2006; Bamberger, Rugh, & Mabry, 2006).

There is, in fact, a range of options to evaluate conservation interventions. Conservation evaluators and managers, however, are largely unaware of these options and lack guidance on how to determine which to use. Although there is no one right way to do an evaluation, there are some general considerations for choosing a design: (1) purpose of the evaluation, (2) program structure and circumstances, (3) resources available for the evaluation, and (4) capacity of those doing the evaluation (Rossi, Freeman, & Lipsey, 1999). Collectively, they influence the extent to which evaluators can use control or comparison groups and the amount, type, and quality of data they can collect. Reality dictates that the type of approach chosen must involve trade-offs on issues such as precision, cost, and buy-in. As such, not all evaluations need or should strive to produce results that establish absolute causality, maximize external validity, and rule out all other explanations (Donaldson & Christie, 2005; Mark & Henry, 2006). The main question evaluators must resolve is whether they have to show that all signs point to an intervention working or whether they need to produce irrefutable scientific proof that the intervention works.

This paper lays out designs available for evaluating conservation actions. It focuses on evaluation for management effectiveness (Stem et al., 2005) and not on status assessment, or methods and tools that are somewhat independent of design and thus can be applied across a range of evaluation designs (Margoluis & Salafsky, 1998). Finally, this chapter attempts

to describe when various evaluation designs should be applied and how approaches may be combined.

### **Design Options for Conservation Evaluations**

Trochim (2006) defines research design as “the glue that holds the research project together. A design is used to structure the research, to show how . . . the samples or groups, measures, treatments or programs, and methods of assignment work together to try to address the central research questions.” This definition easily applies to evaluation design as well. Drawing on this definition, we identify these elements of evaluation design: sampling methods, use (or not) of controls and comparisons, timing of interventions, and timing of observation(s).

As in other fields, conservation evaluation may be based on quantitative or qualitative data (Figure 8.1). Quantitative evaluation can be divided into three broad categories: experimental, quasi-experimental, and nonexperimental (Rossi et al., 1999). Qualitative evaluation relies heavily on purposefully sampling a relatively small number of cases or key informants and rarely uses counterfactuals (Patton, 2002). Qualitative sampling methods are not completely distinct from quantitative evaluation designs in that some qualitative sampling methods can be applied to quasi-experimental and nonexperimental designs (e.g., criterion sampling or stratified purposeful sampling). Nevertheless, we include them in this discussion because they are all evaluation approaches available to conservation practitioners.

Figure 8.1 also discusses internal and external validity, which are key issues in evaluation design. We do not discuss construct validity here because it is largely independent of evaluation design and therefore less relevant to this chapter.

Historically, most conservation evaluations have been purely descriptive in nature, relying heavily on subjective analysis. At best, these evaluations have tended to be qualitative, with little or no deliberate attention to evaluation design. Almost always, they are *ex post* in nature. When conservation evaluations do attempt to infer causality, the most commonly used designs are nonexperimental (i.e., the only sample measured is the one exposed to the conservation intervention).

In this chapter, we briefly discuss some key design options and provide examples of how these options might be best used in conservation. We do not emphasize experimental design here because we believe it is not a particularly cost-beneficial option for real-world conservation projects that wish to conduct management effectiveness evaluation.

**Quantitative Designs.** Quantitative evaluation design includes experimental, quasi-experimental, and nonexperimental designs. These designs allow population-based analysis of an intervention’s impact. In particular, units are identified and sampled so as to allow inference of a population. If a sample has been selected without bias, statistical generalizations can be

### Figure 8.1. Types of Evaluation Design

---

#### Quantitative Design

1. **Experimental:** random assignment of subjects to treated (experimental) and untreated (control) groups

**Advantages:** approximates counterfactual condition; strong evidence for causality

**Limitations:** expensive; often not practical; ethical issues; high expertise

**Validity:**

Internal: high; random assignments; strongest design for internal validity

External: low; artificial setting limits ability to generalize to other settings

**Example:** *Randomized pre and post:* researcher randomly assigns items into control and experimental groups. Measurements taken before and after intervention

2. **Quasi-experimental:** similar to experimental but lacks random assignment

**Advantages:** easier to establish than true experimental designs; fairly strong evidence for causality

**Limitations:** moderately expensive to expensive

**Validity:**

Internal: moderate; inability to randomly assign controls, lack of control over variables

External: moderate; “natural experiments” allow some generalization

**Examples:**

- A. *Matched controls:* intervention group matched with controls selected by researcher
- B. *Regression-discontinuity:* pretest/posttest design in which participants are assigned to program or comparison groups on the basis of a cutoff score on a preprogram measure
- C. *Statistically equated controls:* exposed and unexposed groups or items compared by means of statistical controls
- D. *Generic controls:* exposed group or items compared with outcome measures available on general population

3. **Nonexperimental:** draws inferences about the effect of a treatment on subjects, where assignment of subjects into a treated versus control group is outside the researcher’s control

**Advantages:** least expensive quantitative design; easier to implement

**Limitations:** observe state of world without manipulating it, so less power to detect causal relationships

**Validity:**

Internal: low; no randomization, no controls

External: moderate; natural settings make generalizability stronger

**Examples:**

- A. *Pretest/posttest:* subjects measured before and after intervention

- B. *Time series*: large aggregates taken on large population and compared before and after intervention
- C. *Cross-sectional studies for nonuniform programs*: subjects differentially exposed to intervention compared with statistical controls

### Qualitative Design

4. **Qualitative sampling options**: qualitative evaluation design options focus almost exclusively on the sampling framework and not statistical power or how exposed and nonexposed cases are compared. Individual cases are weighted more heavily because the evaluator is not looking for population-based trends.

**Advantages**: generally, less expensive than experimental and quasi-experimental designs; rich data and anecdotes

**Limitations**: analysis more difficult; subjective interpretations

**Validity**:

Internal: low; no randomization, no controls; researcher interpretation, interviewee perception, recall accuracy

External: low; if cases are carefully selected and analyzed over extended period of time, can be moderate

**Examples** (see Patton, 2002, for more):

- A. *Stratified purposeful sampling*: stratifying samples within samples by selecting particular cases that vary according to a key dimension, thus facilitating comparison
- B. *Extreme or deviant case sampling*: learning from highly unusual manifestations of issue of interest (e.g., outstanding successes and notable failures, top of the class or dropouts)
- C. *Theory-based or operational construct sampling*: sampling subjects on basis of their potential manifestation of a theoretical construct so as to elaborate and examine the construct

made from the sample to the population. Quantitative design is most useful to isolate the effects of specific variables of interest.

*Quasi-Experimental Designs*. Quasi-experimental designs are most appropriate when conservation evaluators want to test the efficacy of a specific conservation tool, cannot take a true experimental approach, and need to have relatively high confidence that the observed effect of the intervention is real. These designs require good knowledge of sampling, instrument and protocol design, and statistics.

One quasi-experimental design option that is virtually nonexistent in conservation but that has enormous potential is matched control/comparison groups. Like other quasi-experimental approaches, the matched controls/comparison group design mimics experimental design, but it lacks random assignment of exposed and comparison groups. Those units that are exposed to a conservation intervention are “matched” with

units that are not exposed. Matching occurs on key variables and controls for potential differences between the two groups, thus isolating the effects of the intervention.

Margoluis et al. (2001) used a matched control/comparison design to determine the utility of sustainable agriculture as a conservation tool. Researchers interviewed farmers who participated in a sustainable agriculture program. Each sustainable agriculture farmer (“user”) family was matched to a farmer family that did not use sustainable agriculture (“nonuser”) techniques. Matched variables included access to goods and services, socioeconomic status, and distance to available agricultural lands. Data collection occurred in two similar biosphere sites in Mexico and Guatemala. Evaluation results demonstrated that users in Mexico planted less area with corn and beans than nonusers, thus reducing the deforestation rates. Conversely, in Guatemala users planted more area with corn and beans than nonusers did, thus increasing the deforestation rate. This design permitted evaluators to isolate the effects of a specific tool and assess the difference between two seemingly similar sites. The three independent variables that predicted the observed differences between Mexico and Guatemala were land tenure status, government land use rights policy, and proximity to protected areas.

*Nonexperimental Designs.* Nonexperimental designs are most appropriate when conservation evaluators want to determine whether there is a relationship between an intervention and an impact, but they do not have the resources to use an experimental or quasi-experimental design, or they do not have access to an adequate control or comparison group. Although results from evaluations employing these designs can be quite useful for management decisions, the designs themselves are generally less rigorous and thus may not require the same level of technical and statistical expertise that quasi-experimental and experimental designs require.

One of the most basic nonexperimental designs involves measuring exposed units before and after an intervention (pretest, posttest, and time series designs). Despite its simplicity and power to infer causality, it is rarely used in conservation. Nevertheless, it is well suited for some common interventions, such as environmental education or media campaigns where one wants to see a change in knowledge, attitudes, or behavior.

A specific example can be found in Powers (2004), where the author evaluated a forestry field visit for second grade students using a pretest-posttest design to assess changes in attitudes and knowledge prior to and following the field experiences. One of the most revealing findings was that socioeconomic variables showed stronger associations with attitude and knowledge changes than did field visit length. The greatest change in knowledge and attitudes was seen in students who participated in the program for only one day but came from communities with less access to forestry settings. Similar environmental education studies using a time-series design are Brossard, Lewenstein, and Bonney (2005) and DiEnno and Hilton (2005).

**Qualitative Design Options.** Qualitative evaluation designs permit in-depth analysis of a particular subject. They commonly focus on knowledge, perceptions, attitudes, and behavior. Typically, qualitative designs are variations of case-study approaches in which key individuals are interviewed as special cases that can elucidate a specific subject. Sampling, therefore, is not population-based and thus does not lend itself to statistical analysis. Moreover, qualitative design rarely involves the use of control or comparison groups. Even so, well-designed case studies can be used for theory development, hypothesis testing, and generalization beyond the individual case (Flyvbjerg 2006). Some common general approaches to qualitative design are participatory appraisal, fourth generation evaluation, and utilization-focused evaluation (Chambers, 1997; Guba & Lincoln, 1989; Patton, 1997).

Qualitative designs include several potential sampling strategies (Patton, 2002), sampling being a main element of an evaluation design. A common qualitative sampling strategy in other fields though rarely used in conservation, is stratified purposeful sampling. Stratified sampling divides a population into relatively discrete groups based on key variables. Purposeful sampling involves selecting a relatively small group of units within each stratum to control for potentially important differences among them.

A hypothetical example of this design involves a conservation and development project to increase household income. Project managers hope to reduce overharvesting of a bamboo species that is an important food for key animal species. The managers believe that providing an alternative to bamboo collection that yields greater financial return on less invested labor will lead community members to harvest less. The evaluators use a wealth ranking tool to analyze community wealth distribution and identify four discrete economic classes or “strata.” For each stratum, the evaluators purposefully identify, select, and interview ten heads-of-household to determine if the project had any influence on their behavior and if this outcome varied across the wealth categories.

This design cannot generate statistical evidence of the intervention’s impact, but it can provide insightful evidence of the difference between groups, and why these differences occur. Results could also inform future quantitative evaluation activities, such as a subsequent household survey, or they could help explain findings from an initial quantitative survey.

## Discussion

The design options presented here hold much promise for conservation evaluation. Quasi-experimental designs generally yield the greatest internal validity while quasi- and nonexperimental designs offer greater external validity and are more precise than qualitative approaches. Qualitative approaches, however, are generally more adaptable and produce more in-depth, revealing information regarding *why* an intervention has worked or not. As such,

depending on the question asked, an evaluator may choose a qualitative or quantitative design.

**What Are Unique Challenges to Evaluation in Conservation?.** The conservation community needs to educate itself about evaluation design options and how different designs produce different results. Nevertheless, there are unique challenges that influence evaluation design selection in conservation. To fully tap into the power of different evaluation designs, the conservation community must address these challenges, and in many ways change how it operates. Among these challenges:

- *The units acted on are often not the units conservation projects want to ultimately influence.* A public health project might provide immunizations to children in a community. The intervention is on children, and the ultimate impact is measured in terms of the children's health. In contrast, conservation projects are often designed to influence individuals, governments, or societies, but their impact is measured in terms of species and ecosystem health. Not surprisingly, extrapolating from an intervention aimed at humans to an impact on species or ecosystems is complex and challenging.
- *Conservation projects usually take place in large, complex settings.* Whether a large site such as the Amazon Basin or a global theme such as climate change, conservation initiatives work in complex, dynamic, and often unpredictable circumstances.
- *Relevant secondary data are rarely available.* Conservation project evaluation requires biological and biophysical data for the dependent variables. Although many conservation researchers and organizations collect this type of data, there is often little coordination between their agendas and those of project teams. Consequently, if secondary data do exist for a project site, they are often not the right data or at the right scale to answer management questions. Conservation organizations therefore must collect primary data for evaluation purposes. This is often costly and difficult to do.
- *Evaluation is rarely incorporated into program design.* Most conservation evaluations are an afterthought with no attention to evaluation requirements at a program's inception. As a result, no baseline data are collected, and little or no data are collected throughout the project's life.
- *Conservation planning and evaluation require knowledge of and expertise in social and biological sciences.* Recently, the conservation community has been somewhat divided into "biologist" and "social scientist" camps and has done little to integrate social and biological data. This integration is crucial because ultimate impact is measured in terms of species and ecosystem health while intermediate outcomes are typically measured in terms of human knowledge, attitudes, and behavior.
- *Conceptual linkages among interventions, outcomes, and impacts are relatively unknown.* Because conservation is a relatively new field, many of the conceptual underpinnings that delineate associations and potential causality among interventions, outcomes, and impacts are untested and



unknown. Designing conservation evaluations at the outset, therefore, is especially challenging because data requirements are unclear and expected results are unpredictable.

- The *time lag between intervention and impact is often very long*. Although not exclusive to conservation, it is especially problematic in this field. Because of life cycles and natural fluctuations in species and ecosystems that occur over many decades, changes attributable to a particular intervention are difficult to establish.
- *Urgency drives program planning more than evidence of success does*. The conservation field is one driven by perceived urgency to act; thus there has historically been little will to “divert” precious program dollars to evaluation activities. Consequently, the field has not been too systematic in learning about which interventions work best under which conditions.

### **What Can the Conservation Field Do to Address These Challenges?**

To overcome these challenges and be better prepared to choose the right design, we suggest a number of evaluation-related principles.

*Build Appropriate Evaluation Design Into Program Design.* Conservation practitioners and evaluators must consider expected outcomes at the outset—not only at the end—of a project, and they must collect baseline data. They must also understand the conditions under which specific evaluation designs are appropriate. In particular, we call for an end to the false dichotomy that pits the use of counterfactuals against all other forms of evaluation. The question is not, “Which is the best evaluation design in conservation?” Instead, it should be, “What is the best evaluation design option available to conservation managers for their circumstances?”

*Ensure an Adequate Theory of Change.* Most conservationists erroneously believe that by measuring only the dependent variable (e.g., species and ecosystems), one can tell if conservation interventions have been successful. In fact, to reach this conclusion one must measure incremental change at various points along a theory of change, from the intervention to intermediate outcomes to ultimate impacts (Foundations of Success, 2007). What interventions led to knowledge change? What knowledge change led to attitude change? What attitude change led to behavior change? What behavior change led to a mitigation of threats? Which threat mitigation led to species or ecosystem improvement? These “incremental evaluations” help evaluators break down complexity, understand system components, and reconstruct an understanding of the conditions within which interventions operate. They also foster incremental learning and help develop a body of evidence—a plausible case for association, causality, and ultimately conservation impact.

*Partner With Organizations That Have the Expertise to Collect Social Data.* Until the conservation community builds its internal capacity to effectively measure the human dimensions of conservation, it should seek partners in relevant fields to identify appropriate indicators, collect and analyze data, and link changes in human dimensions to changes in biological dimensions.

*Bridge Biologist-Social Scientist and Practitioner-Researcher Gaps.* Conservation organizations should add to their staff individuals with training and experience in bridging the social and biological divides who are capable of playing both programmatic and evaluative roles. These individuals can help create a robust subfield of operations research (as in public health) that accelerates learning and optimizes effectiveness. Likewise, conservation organizations should work closely with scientists to coordinate research agendas to help answer important management questions.

*Do Not Sacrifice What Is Important for What Is Urgent.* If the conservation community is to make concrete advances in understanding which tools and strategies work best under which conditions, it must stop, look, listen, analyze, and learn; that is, it must vastly improve its ability to evaluate actions. To do this, it should invest more heavily in its capacity to design and implement effective evaluations.

**How Can Design Alternatives Help the Conservation Community Overcome These Challenges?** The evaluation design option that is most appropriate to a given conservation situation depends on the purpose of the

**Table 8.1. Conservation Evaluation Scenarios and Potential Designs**

<i>Scenario</i>	<i>Example of Appropriate Design or Approach</i>
A large conservation organization wishes to gauge the effectiveness of its policy reform campaign to change greenhouse gas emission laws.	Qualitative purposeful sampling in which members of Congress are interviewed
A foundation wants to see how effective its partner capacity building program is.	Nonexperimental, time series in which participants are interviewed three times before and three times after participating in a training program
A government agency with limited evaluation capacity wants to know whether an increase in park fees has improved park conservation.	Mixed method approach: 1. Review of secondary data (accounting books) 2. Qualitative purposeful sampling in which park rangers are interviewed to determine their perceptions regarding changes in animal populations
A regional marine conservation organization wants to gauge the impact of its “best management practices” program for fishermen.	Quasi-experimental matching in which fishermen who use new technologies are compared to those who do not
A national conservation organization wishes to measure the success of its media campaign to buy certified wood.	Nonexperimental, pretest/posttest with a large sample in which consumers nationwide are interviewed on exiting a national home improvement store
A small conservation organization with limited resources wants to know if its income generation project is working.	Qualitative purposeful sampling in which four age classes of women participating in the project are compared

evaluation, program structure and circumstances, resources available, and the capacity of those doing the evaluation (Rossi et al., 1999). One tactic to draw on the merits of both qualitative and quantitative design is to use a mixed-method approach. Mixed methods offer a variety of benefits (Patton, 2002; Bamberger et al., 2006; Greene, Kreider, & Mayer, 2005) to strengthen conservation evaluations. The examples in Table 8.1 illustrate a range of potential options for how conservation evaluators could use different evaluation designs and approaches under different conditions.

## Conclusions

Evaluation is not a one-size-fits-all endeavor. The “best” evaluation design for a project will vary widely by factors including available time and funding, existing expertise, and level of precision needed. Trade-offs are inevitable. Does one invest large sums of money over a long time to establish causality with certainty? Or does one seize an opportunity to learn quickly and adapt in a dynamic, changing world? Less rigorous evaluation designs mean there will be situations where evaluators and practitioners miss causal links and make the wrong decision. Good, continuous, systematic monitoring, however, is crucial because it allows practitioners to tweak their actions as they implement, catching poor decisions early. Unfortunately, this type of monitoring and adapting is all too rare in conservation.

We have briefly described the range of options available to evaluate conservation project effectiveness and some design considerations that are somewhat unique to conservation. We also have attempted to show how they can be used effectively by themselves or in combination for robust evaluation findings.

Evaluation in conservation is not only about measuring effectiveness. It is also about engaging the conservation community. If we disassociate evaluation from practice, we will lose the practitioners and as a result their commitment to using evaluation results. To ensure evaluation leads to management decisions, conservation managers must engage in choosing the right evaluation design. More generally, the conservation community needs to discuss which evaluation designs work for its practitioners and under what circumstances they are most appropriate. With this chapter, we hope to take an initial step in that direction.

## References

- Bamberger, M., Rugh, J., & Mabry, L. (2006). *Real world evaluation: Working under budget, time, data and political constraints*. Thousand Oaks, CA: Sage.
- Brossard, D., Lewenstein, B., & Bonney, R. (2005). Scientific knowledge and attitude change: The impact of a citizen science project. *International Journal of Science Education*, 27(9), 1099–1121.

- Chambers, R. (1997). *Whose reality counts? Putting the first last*. London: Intermediate Technology.
- DiEnno, C. M., & Hilton, S. C. (2005). High school students' knowledge, attitudes, and levels of enjoyment of an environmental education unit on nonnative plants. *Journal of Environmental Education*, 37(1), 13–25.
- Donaldson, S. I., & Christie, C. A. (2005). The 2004 Claremont Debate: Lipsey vs. Scriven—Determining causality in program evaluation and applied research: Should experimental evidence be the gold standard? *Journal of Multidisciplinary Evaluation*, 3, 60–77.
- Ferraro, P. J., & Pattanayak, S. K. (2006). Money for nothing? A call for empirical evaluation of biodiversity conservation investments. *PLoS Biology*, 4(4), 105.
- Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry*, 2(2), 219–245.
- Foundations of Success. (2007). *Using results chains to improve strategy effectiveness*. Bethesda, MD: Author.
- Greene, J. C., Kreider, H., & Mayer, E. (2005). Combining qualitative and quantitative methods in social inquiry. In B. Somekh & C. Lewin (Eds.), *Research methods in the social sciences* (pp. 274–281). London: Sage.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.
- Margoluis, R., Russell, V., Gonzalez, M., Rojas, O., Magdaleno, J., Madrid, G., et al. (2001). *Maximum yield? Sustainable agriculture as a tool for conservation*. Washington, DC: Biodiversity Support Program.
- Margoluis, R., & Salafsky, N. (1998). *Measures of success: Designing, managing, and monitoring conservation and development projects*. Washington, DC: Island Press.
- Mark, M., & Henry, G. (2006). Methods for policy-making and knowledge development evaluations. In I. Shaw, J. Greene, & M. Mark (Eds.), *Handbook of evaluation: Policies, programs and practices* (pp. 317–339). Thousand Oaks, CA: Sage.
- Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs*. San Francisco: Jossey-Bass.
- Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Thousand Oaks, CA: Sage.
- Patton, M. Q. (2002). *Qualitative evaluation and research methods* (3rd ed.). Thousand Oaks, CA: Sage.
- Powers, A. L. (2004). Evaluation of one- and two-day forestry field programs for elementary school children. *Applied Environmental Education and Communication*, 3(1), 39–46.
- Rossi, P. H., Freeman, H. E., & Lipsey, M. W. (1999). *Evaluation: A systematic approach* (6th ed.). Thousand Oaks, CA: Sage.
- Stem, C., Margoluis, R., Salafsky, N., & Brown, M. (2005). Monitoring and evaluation in conservation: A review of trends and approaches. *Conservation Biology*, 19(2), 295–309.
- Trochim, W. M. (2006). *The research methods knowledge base* (2nd ed.). Retrieved version current as of October 20, 2006, from <http://www.socialresearchmethods.net/kb/>

RICHARD MARGOLUIS, CAROLINE STEM, NICK SALAFSKY, and MARCIA BROWN are affiliated with Foundations of Success (FOS) in Bethesda, Maryland.