E v a l u a t i n g E v a l u a t i o n

for Wildlife Conservation Society's
Conservation Science Training Course and
Small Grants Programs

in Venezuela and Colombia

by Jon Kohl
Yale School of Forestry and Environmental Studies
New Haven, Connecticut
April 1997

Evaluating Evaluation

The report defines and illustrates three types of evaluation that can be applied to the case programs. From these WCS managers can adapt the options according to their purposes.

Valuing Evaluation

Exploring Evaluation Diversity

Evaluations are as diverse as rainforests. They can be simple or complex. Some evaluations begin on the first day of project planning, others begin only after years of operation. They critique winning programs and losing programs. Sometimes they increase prestige, improve functioning, or only raise funds. They can be done by outside experts or can be integrated into staff responsibilities. They can transpire in quiet times or politically volatile times. Some are information intensive, others less so. Some are rigid and top-down, others adaptive and participatory. An evaluation can ravish staff morale or build up self-confidence. They may be generally desired or can evoke a conflagration among stakeholders like the butting heads of a mythological hydra.

Setting Sights on WCS Programs

Many eyes view an evaluation, often from very different perspectives. This report speaks to some of these differences and offers WCS several suggestions on how to go about using an evaluation. WCS's Strengthening Manpower Program whose small grants program and conservation science courses have used heretofore an informal reputability assessment to judge success, that is, a collection of opinions of various stakeholders: program staff, alumni, other funders, etc. While in many arenas reputability proves sufficient, more can be done. In this age of increasing accountability and dwindling funds, this report embarks upon a different approach that will consider those personal judgments, and will go beyond.

This report thus does three things for WCS program staff:

1) It conducts a partial evaluability assessment and demonstrates how to proceed.
2) It explains an implementation evaluation and gives examples.
3) It suggests how to design short- and long-term outcomes analyses and offers an exemplary alumni survey instrument (appendix).

The overall goal of the report is to give program managers a wide understanding of some of the major choices in evaluation and offer options they may use to proceed in an evaluation process. The report itself then seems a kind of Frankenstein of evaluations, part evaluability assessment, part implementation evaluation, and part outcomes analysis. This body of assorted parts represents not so much confusion but rather a synthesis of function and available data. I have interviewed the WCS program coordinator and read through a variety of documents both internal and promotional. From these I have been able to uncover the bones of the programs sufficient to offer several options for evaluating the programs. The job cannot be completed, however, because much information remains to be collected (most of which is in-country) and many decisions remain as of yet, unmade. Some sources of information to be examined might include:

1) Interviews with in-country program staff and other stakeholders
2) Use of program records
3) Survey of participants/alumni
4) Records of other institutions like universities

5) Survey of control groups
6) Interviews with key informants at universities, funding agencies, research institutions
7) Observation of the programs in action

This report shows how evaluation can be used effectively for these two programs, by beginning each of three kinds of evaluation, it demonstrates their utility. To guide this process of discovery I socratically ask key questions throughout. This process is forthright, meant to challenge not criticize, and thus is an internal document intended to be reshaped to the various purposes that WCS puts it. I have written with the understanding that an evaluation should be the servant of social policy, not its master. And above all, an evaluation's quality should be determined not by the answers it generates, but by the questions it asks. (See box 3 at end of document.)

Evaluating Evaluability

Mapping the Social Terrain

An evaluability assessment could also be called a pre-evaluation assessment and aims to assess or map the "social reality" of program operation as opposed to the "written reality." Understanding this social reality includes learning how the program operates, what the stakeholders think, and identifying what political or resource wants may limit evaluation. This qualitative approach functions also to facilitate agreement among the stakeholders on the need for an evaluation (if it is needed), on what kind of evaluation it should be, and on who will be involved and what their part in the process will be. The participation of program staff in this early step integrates them into the process and lets them know they are an important component to this program. Their participation in effect aims to achieve their collaboration ("buy in") and agreement in the evaluation process. It also trains them to carry on the evaluation internally in the future. Some points of agreement include:

1) Program objectives
2) Program components to be analyzed and kind of evaluation appropriate
3) Commitment of resources and cooperation/collaboration
4) Plan for using results
5) Plan for efforts required to make the programs more evaluable and how they can use this plan later to evaluate the program in house

Staking Out the Stakeholders

Every program has numerous stakeholders whose numbers tend to increase as the project ages. Different stakeholders socially construct problems in different ways and those ways can seriously affect the entire evaluation process. Many evaluations fall flat on their face because stakeholders do not cooperate in the process or the results are ignored. This often happens because different, and, at times, competing interests of stakeholders were never reconciled and the evaluation never earned the cooperation and interest of those parties it desperately needed to avail. Thus an evaluability assessment helps to foment discussion among stakeholders and determine the evaluation's parameters.

For example I am one stakeholder in this evaluation. My interests are to discuss and apply a range of approaches I am studying and making as honest and rigorous attempt at

evaluating the given programs which may lead to potential work sometime in the future. The interest of the WCS program coordinator may be to produce an evaluation whose first goal is to raise funds and second to offer helpful suggestions. Thus at times these two interests may bump each other. Such cross-purposes permeates many evaluations.

These programs in fact have many other stakeholders as well, including the WCS development, administrative, and public relations offices; it has in-country managers; international and national funding agencies; universities; program participants; research institutes such as the Alexander von Humboldt Institute in Colombia; federal agencies; and a host of local collaborating institutions — all of which could potentially precipitate disagreements.

Each one of these stakeholders to some extent has a different set of interests and thus could derive different utility from an evaluation. Not all their interests have to be integrated into the design of the evaluation but a complete evaluability assessment would consider their perspectives and try to anticipate the effect of their influence on the evaluation.

Holding an Evaluability Workshop

Stakeholders must agree on the purpose and the parameters of the evaluation which include problem definition, impact models, operationalized objectives, time frame, expectations of cooperation, and funding. This can be accomplished through an in-country workshop between program staff, evaluators, and other key stakeholders such as alumni. By consensus the forum should write up a Statement of Work which should describe the role of the evaluation, include the goals, composition of evaluation team, responsibilities of all involved parties, and time lines of deliverables. A workshop can be a key step in achieving staff "buy in" to the evaluation.

It is important to remember that the evaluation could affect the staff in a variety of ways. The achievements and concerns are perceived, interpreted, and communicated through an evaluator's beliefs, values, and assumptions. Evaluators are in a position to have a profound impact on projects and participants. They can build or destroy confidence. They can enhance or undermine social and political bridges. Thus it is important to carry out this phase with sensitivity.

Evaluating the Implementation

Ideally an implementation evaluation compares the ideal design of the program and that which really exists. The differences between the two designs reveal problems encountered by the implementation. It also reveals innovations in implementation that prove more effective than the original design. The design of a program of course follows from its operationalized objectives as discussed below. Since no original project design exists in this case, the first step to evaluating the implementation is to describe what is going on.

Describing the Program

Having a program description or blueprint of operation is essential for evaluating the efficiency and efficacy of an implementation. It is also necessary if the program has any hope of expanding or being replicated into new areas. Having a plan increases a program's credibility in the eyes of donors and most of all tells evaluators how the program is supposed to affect program participants. It happens, though, that evaluators often have to construct a program description for the first time since no formal structure of the program had ever been put

down on paper. This is the case with these programs.

Since a program description requires observation of the program and discussion with the implementing personnel, such a task cannot be done here (or even in this country) except with the broadest of strokes. Nonetheless I list areas such an evaluation would cover should WCS decide to undertake one.

Treatments and services

In order to be evaluable, we must ask what are the treatments? How are the treatments applied (in full or partially)? Are the treatments applied consistently across participants? For the small grants program, we know that applicants receive advice, guidance, and some are "discovered" by program personnel, presumably to be groomed for their potential. This personalized treatment, if true, however, ravages the possibility of comparing different treatment groups as the treatments are unevenly applied.

Qualifications and competencies of staff

Who are the staff members, what are they trained to do, and what responsibilities do they have? Are staff competencies adequately utilized (under/over) by the program? Who manages the budget? Certainly different treatment require different competencies. The small grants program, aside from a typical administrator, requires also a respected scientific selection committee to judge the applications and to make funding decisions. Their qualifications affect the program's credibility which affects the application rate whose degree of exclusivity (measure of prestige) affects the treatment. Prestige affects the psychology and self-perception of the grantees themselves versus control groups.

Mechanism for recruiting and obtaining cooperation of targets

Do program materials actually reach target population and reach them evenly across the population? In 1995 Fundación FES of Colombia mailed out brochures and posters to over 300 institutions nationwide in its search for applicants. It is quite possible through this distribution pattern, or through brochure design, or a series of other factors, applicants are not reached or not persuaded to apply. The procedure for application itself may pose systematic barriers to certain kinds of applicants who live farther away or work with less known universities or for whatever reason.

Means of optimizing access to program

It may be that applicants are required to pick up the funds personally and the building is difficult to locate. This is more an issue for the conservation science course. Are applicants dissuaded from participating due to the course's distance? Does housing become a prohibitive expense? Are there other barriers to program access for certain groups of the target population?

Target retention, referral, follow-up

What is the attrition rate of the programs? Why and who drops out of the programs? What efforts do the programs take to reduce attrition and how well does the program keep up with

graduates after the termination of their program? How is program completion defined? (Delivery of a final report and graduation constitute termination of the small grants program and conservation science course respectively.) Referral applies more to other social programs where participants seek services that the program does not offer and need to be referred elsewhere. If the small grants program, however, explicitly attempts to advise and mentor, this might involve referring students to different experts in other field. This might be a formal part of the networking goal of the program.

Activities of the program

What are the activities of the staff? Of the program? Where do they take place? How are they carried out? What equipment is used? How much does it cost? What is the annual cycle? How are applicants selected? The activities form the heart of a program and quite often the cause of inefficiencies. A careful documenting of the activities is essential for replication and for rooting out certain kinds of procedural problems. Though anecdotal, I have far more quarterly reports from the Program Administrator in Colombia than in Venezuela. Why is there such a discrepancy in the reporting activity to WCS?

Making the Parts Conform

The definitions of the problem, goals, operationalized objectives, and impact models constitute the first determinants in the implementation of a program as well as being the building blocks of any program design. Often a program may be stricken from the outset if different stakeholders define problems differently. Different problem definitions precipitate entirely different programs. Conversely a poorly defined problem results in vague objectives and a program design that may not relate at all to the original purpose of the program. While I offer suggestions for definitions, objectives, etc. they are meant to carry forth the demonstration of how evaluation can be used. They may be kept or discarded, because the questions are most important.

Problem Definition

What is the problem? Can it be defined specifically and to the agreement of stakeholders? First we look to the documents to find articulation of the problem definition. From the letter of intent to the Moriah Fund (June 1996), WCS writes that

> there is a dearth of conservation professionals sufficiently trained to monitor and manage their country's natural resources. Thus, a cornerstone of WCS's conservation program is building professional capacity — mentoring, advising, and enhancing the skills... We do this through formal courses, as well as through fieldwork where 'students' are taught survey and monitoring methods, data analysis and interpretation, and most importantly, how to apply that data to the problems at hand. Possibly one of the most valuable aspects of our training activities is the small grants program: giving young professionals the opportunity to carry out their own applied research and conservation activities.
>
> In Venezuela, conservation professionals, until recently, were virtually non-existent. Although, CONICIT, an NSF-like institution, provided some research opportunities for established researchers and university professors, there were few opportunities for budding professionals who were finishing school and trying to establish their careers.

Jessie Noyes Foundation proposal (1989): "In the long term, of course, WCS's goals is to

increase capacity of Latin American countries to value and protect natural resources, in particular the rainforests as sites of greatest biological diversity and long-term economic value."

Of course, this problem could be fleshed out to be more useful. The problem is not that there are not enough trained natural resource managers, but that the dearth of them results in

•increased degradation of natural resources;
•less political pressure to pass environmental laws;
•international embarrassment;
•high costs in hiring foreign consultants to do the work;
•lower level of protected area establishment because there are insufficient qualified managers;
•less funding for biology and conservation research programs because there are not enough qualified candidates.

Or we could step back and say that the lack of qualified personnel is not the problem, it is the effect. The problem may be that

•there are insufficient funds to promote training programs;
•the universities are not equipped to train this kind of personnel;
•the government which pays the salaries of protected area managers and scientists regards the issue of such low priority that salaries cannot compete with other opportunities such as in the private sector and thus candidates do not enter the field.

Even if these factors were clear that would describe the *nature* of the problem. But what of its *distribution*? Is this problem relegated only to

•women rather than men?
•whites and mestizos rather than indigenous or blacks?
•poor rather than rich?
•Amazonian areas rather than Andean areas?
•the 1980s rather than 1990s?

The social context is dynamic and factors change. But once we have an understanding of the full context of the problem, it may become clear that one kind of program is more effective than another. If there is no money to hire more qualified personnel, then training them will only exacerbate the problem. Or if politics is responsible for the problem, perhaps a communications campaign targeted at decision-makers would be more appropriate than a training program.

For the purposes of this discussion we will define the problem as follows (adding fictional elements where necessary):

Due to an historical lack of investment in the science and university systems (almost all public), very few opportunities existed to train students for conservation careers. The lack of such skilled personnel has resulted in an underestablishment of protected areas (presumably limited by science, not politics) and a backlog of unprepared management plans. This has led to encroachment on and conversion of wild lands at a greater pace than would otherwise happen were there sufficient established protected areas and management plans.

Given that there is no official document describing program operation and also that the sporadic nature with which available documents address problem definition and objectives, it has not been made clear that either program design effectively emanates from amain problem.

For the conservation science course, moreover, an additional problem must be considered. An apparent fatalistic, pessimistic attitude inflicts many people that the conservation problem is so grave that many environmental problems cannot be solved. This sentiment demoralizes students and young professionals alike. The problem derives from university-level education where conservation biology is poorly taught if at all (thus students are not exposed to successful techniques) and is reinforced by mass media negative coverage (including the perception of an unresponsive government), and unpleasant personal experiences. Thus from the perspective of students, conservation can at best be only a short-lived and poorly funded career. This cumulative pessimism conservationists dub the "Syndrome of Learned Hopelessness" (*Síndrome de la Desesperación Aprendida*).

This syndrome combines with the above described lack of training to justify the conservation science training program.

Goals and Objectives

Goals are statements of desired conditions to be realized through a program intervention. Objectives are operationalized goals, explicitly detailing the parameters of the desired outcomes. Throughout the literature of these programs goals have been expressed variously and vaguely. They have not been operationalized in any quantitative sense, and sometimes do not refer back to a major problem. Consider the following examples.

• WCS hosted a regional meeting of its South American affiliated programs in 1996. From the internal report of this meeting, the following section talks about WCS's training effort in Latin America. There they described the large number of training programs. It says,

> Although the program in its diversity is considered a success, WCS's training has not yet gone through an evaluation process and does not have indicators of success. The group reached some conclusions about the future of the training program: training professionals is less risky than training undergraduates because professionals have already decided their career. An emphasis should be made to transform scientists into conservationists... Participants concluded that training is the only way to guarantee sound conservation in the long-term. Furthermore, the group vowed to expand a training program that increases South-South connections, and provide locally-proven solutions within similar 'country contexts.'

Why training is the best guarantee of sound conservation the report does not say. Perhaps it means conservation "science" rather than conservation practice. Perhaps training is easier than being involved in the policy arena (most of the participants were biologists). Perhaps the networking which training promotes serves scientific prestige and collaboration. Without a thorough stakeholder analysis and in-country interviews, it is difficult to pry apart objectives that derive from the principal problem and objectives that derive from other goals of stakeholders.

Also from an earlier part of that report, the goal is talked about generally as building leadership whatever they exactly means. "A consistent and expanding effort during the last 10 years has created a second and third generation of conservation leaders. Training efforts have yielded a wide variety of initiatives that are part of a training 'pyramid' towards the creation of conservation leadership at many levels."

•The Moriah Fund letter of intent mentions the sole goal of the small grants program as "building professional capacity."

•A recent brochure for the program speaks only of "driving the protection, conservation, and sustained use of the enormous biological diversity of [Colombia]."

•In the original manual for the conservation science course it states that the primary objective of the course is to infuse a good dose of optimism (in its students), achieve an open mind in order to understand other points of view, to demonstrate that conservation problems are worldwide and not just localized where students live, and to deliver the message that conservation can be achieved by individuals.

•The grant proposal to the Jessie Smith Noyes Foundation in 1989 (which was funded) says the program aims to make up for a lack of funding opportunities for students in conservation research. Also "university courses provide little formal training in applying conservation biology in the field to achieve lasting protection for threatened habitats and ecosystems." Those are two problems. See Box 1 for a more detailed critique of these goals.

Box 1 Critique of Goals and Objectives of Jessie Noyes Proposal

Goals and Objectives

1. To provide small grants enabling undergraduates, graduate students, and recently qualified conservation scientists in the neotropics to conduct their own independent field projects. Since WCS perceives this field experience as part of the on-going training process, we also aim to provide expert guidance and supervision to grantees.

2. To provide a framework in which individual careers can flourish. Whereas North American and European conservation scientists have well-established infrastructures in which to develop their careers, the opportunities for their counterparts in Latin America are still somewhat less common. It is therefore important for WCS to provide opportunities for individual scientists to become part of a larger regional conservation network, and staff such as Strahl to discover the individuals, help them with their early field projects, and provide the broader context for stimulating and effective careers in conservation. As the recent workshops organized by Dr. Strahl in Caracas have shown, the proposed national and international shortcourses would strengthen this by serving as a forum where issues of regional concern can be addressed and international networks be established.

3. To work with and strengthen non-governmental conservation organizations in the host country. As mentioned above, the new Venezuelan NGO EcoNatura was established in 1989 to handle the first Noyes/WCS grants to Venezuelan students and to serve as long-term national base for conservation training in the future...

****

Goal 1 is the simple execution of the program itself, nothing to do with any positive effect on the conservation of biological diversity. It adds also the provision of expert guidance and supervision. These terms need to be de scribed in order to know what treatment grantees are receiving and to deter mine if it is consistently applied across all participants (this is a delivery is sue).

Goal 2 uses many ambiguous and vague terms: framework in which careers can flourish; well-established infrastructure; develop their careers; dis cover individuals; help; provide broader context for stimulating and effective careers; strengthen this. Noyes cannot know what this goal seeks to achieve.

Goal 3 is understandable but offers no guidelines as to how it will strengthen organizations; instead it gives an example of creating a new orga nization, presumably because a local organization could not be strengthened sufficiently to administer the program. What does a long-term national base for conservation training do?

Impact Models

There are many goals from which to choose and other morphs still hide in the literature. But before one can go about operationalizing objectives, an impact model must underlie its selection. That is, some

set of hypotheses should plausibly explain how the desired change comes about *in vivo*, without the proposed intervention. This understanding informs how the intervention will effect the change. What are the assumptions to the unspoken underlying model? Are there empirical data to support the model? Without a model, even if the program somehow achieves the outcomes, there is no basis for expanding the program to other situations under different conditions. Also just because the change comes about *in vivo* does not mean it can be induced under intervention conditions.

A model should have three constituent hypotheses all of which continually refer back to the principal problem definition and goals.

1) Causal hypothesis: what causes the problem
2) Intervention hypothesis: how will the intervention affect those causes in the causal hypothesis to effect the desired change
3) Action hypothesis: how does the intervention relate to outcomes

Because the problem definition had been weak, even if the programs had tried to develop a model, it would have been quite difficult except possibly for the Syndrome of Learned Hopelessness (SLH). In this case, the program attempted to develop a psychological model based on anecdotal evidence. A psychology literature search could strengthen this model.

To use the SLH as an example, let us pursue it a bit. It argues that due to poor teaching in the university (a pedagogy lacking both factually accurate material and optimistic, successful examples), widespread negative media coverage, a feeling of isolation amidst large-scale environmental problems, prevalent negative personal experiences, and a recognition that both funds and government attention lack, students have developed the so-called Syndrome of Learned Hopelessness. True gloom and doom. Even though the theory is based on anecdotes, such a belief is credible and readily observable.

1) Causal hypothesis: The above factors combine to effect the pessimistic attitude.
2) Intervention hypothesis: By presenting successful techniques, factually accurate material, facilitating communication between students and successful practitioners, demonstrating the global nature of environmental problems, illustrating individuals who have accomplished environmental achievements, and framing the entire class in a positive atmosphere, SLH can be overcome.
3) Action hypothesis: By rolling back the syndrome through the intervention at the period in their lives when they are making career choices (before and after graduation), students should be motivated, sufficiently knowledgeable, and have the intention to pursue and create careers in conservation.

It should be noted that although this model has not been substantiated with a literature search, often times it is not feasible or practical (due to information or time costs) to conduct such a search. Where possible, nevertheless, it is advisable. See Figure 1 for a completely home-made schematic of this model.

Operationalizing Objectives

My interview with the program coordinator revealed the following goals for both small grants and conservation science course:

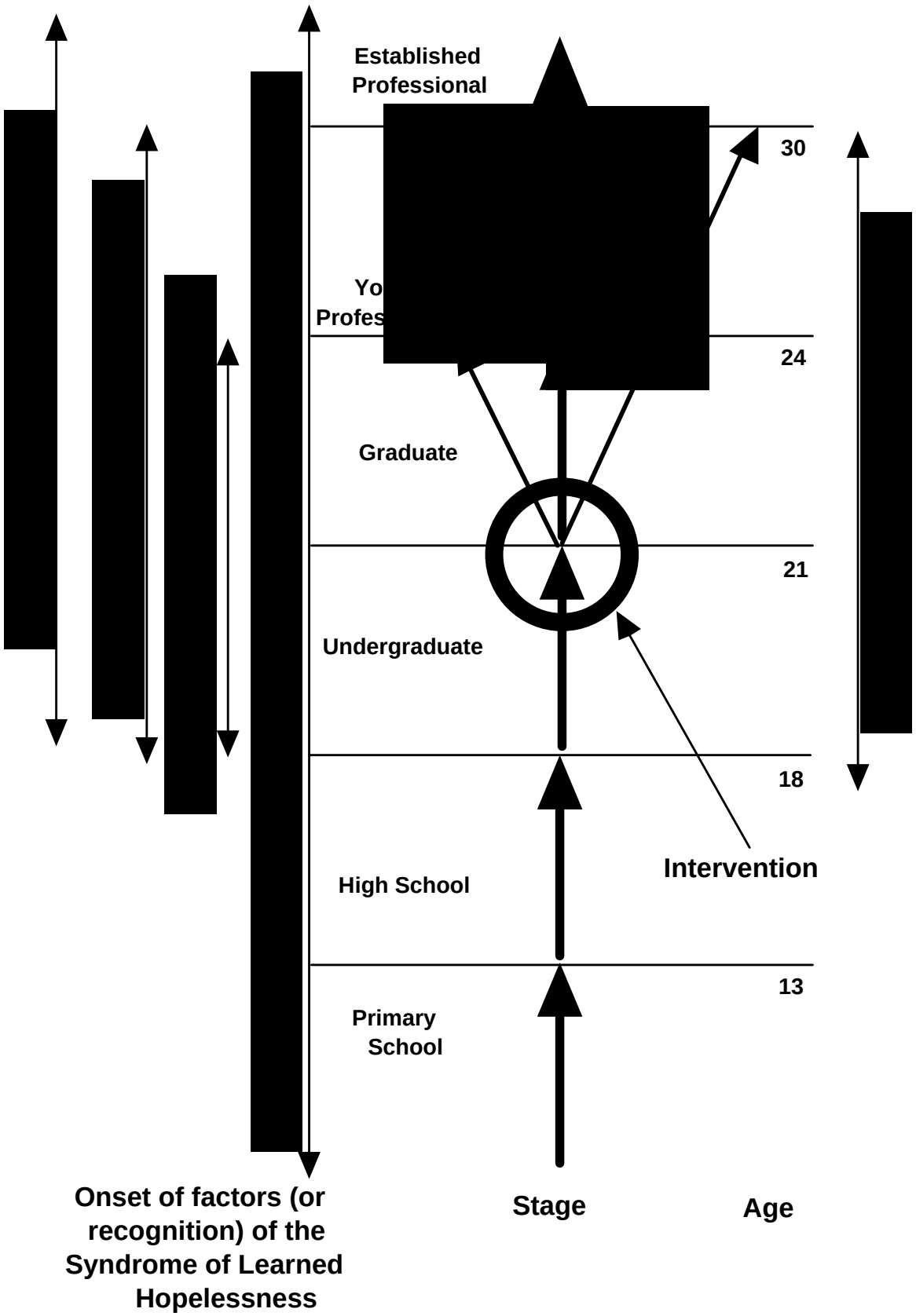

Figure 1 Impact Model for the Syndrome of Learned Hopelessness

1) To get more people to enter the conservation field.

2) To get people to create conservation jobs that did not exist before.

Now let us try to clarify and operationalize these two goals for the conservation science course which will be used hereafter, using a number of questions. See Box 2.

How does one define entering a field?
A graduate must enter and remain in a conservation-related job for a minimum of two years to qualify as entrance. Internships, volunteerships, and jobs of less than two years do not qualify.

More people to enter the field than what?
At least 20% more graduates must enter a conservation-related job than a control group. The difference must be statistically significant.

How about rate of entry?
Graduates of the course must enter the field statistically earlier than members of the control group.

Who defines a conservation job?
WCS will construct a comprehensive list of kinds of institutions that qualify for the category keeping in mind that the problem definition implies managing natural resources in a protected area. Thus a person working in an electric utility is not enough information to be ranked. All stakeholders should review and agree with the categorization.

What does create a conservation job mean?
The funds that pay that job cannot have been transferred some other conservation fund; they must originate from a source that had not directly funded a conservation job. The determination must be made by a panel of experts from several stakeholder groups who are familiar with public and private sectors in the given country.

Box 2 Clarifying Objectives

Each final objective should be written to contain one purpose and one action as well as a specified time frame. The one-and-one approach makes it much easier to determine if the objective has been met. If an objective contains two results and one has not been met, to what extent can the program claim to have met the objective? Thus,

1. To motivate students to enter the conservation field earlier than controls. (Controls will be discussed on page 21.)

2. To motivate students to enter the conservation field at a rate of at least 20% more than controls. (The percentage is arbitrary.)

3. To motivate students to create their own conservation jobs at a rate of 30% more than controls within 10 years of graduating from the program.

Also the following objectives comes from the model.

3. To overcome SLH by having students score statistically higher than a pre-test in
   a. -post-attitudinal,
   b. -knowledge, and
   c. -intention tests taken immediately after the course.

4. To overcome SLH by having students score statistically higher than a pre-test in
   a. -post-attitudinal,
   b. -knowledge, and
   c. -intention test one year after the course.

5. To demonstrate the dissipation of SLH by having students attempt to create jobs at the same frequency as their post-course intention scores predicted.

Covering Coverage

Defining the Target Population

Both programs have primarily targeted undergraduates, graduates, and recently graduated young professionals. On separate occasions the courses have also been attended by high school students and university professors and other advanced professionals. The model posits that the intervention should take place around university level both when students are making career choices and when they are most actively learning hopelessness. Also this is presumably when they first encounter the need to fund raise and be funded. Neither high school students nor university professors meet these criteria. While courses can be designed for them as well, the focus on attitude change would unlikely be the focus of the course design. Certainly any university professor that decided to enter a conservation science course like this (presumably coming from outside the conservation field) would be highly motivated to add conservation techniques to their discipline. It is not likely they are pessimistic about the potential of effecting positive change or about funding if they are already university researchers, and less likely that they will be inspired by successful practitioners since they themselves have already achieved that rank and most likely appreciate the role of individual achievement.

Last high school students still must go through four years of college after the intervention which the model does not attempt to explain. It is highly conceivable that the message of the course would be diluted by the discovery and rapid intellectual growth of four or more years of university study.

Therefore with the exception of established professionals and high school students, the overall target population seems to be correct. As will be seen later, program administrators can select for sub-populations depending on different demographic and disciplinary features that they deem appropriate, e.g., more females, more social scientists, more candidates from distant states.

Too Much, Too Little

Although discussed before, coverage is a major point of success or failure. We have already seen that the program overcovers in some areas, including professors and high school students. How does the program define a young professional? In the model we identified an age, but is the program that discriminating? A study of applicants' records will reveal coverage patterns. These patterns then will be compared with the ideal population and either differences need to be justified or program parameters need to be changed.

If there is overcoverage, the program likely wastes resources on non-target people. If undercoverage, then the program misses part of its target population and thus does not fully achieve its objectives. As time passes, the program may discover new needs which it can formally incorporate into its selection process. In fact, we can consider some of the variables over which the program has control for the small grants program. In the recent Colombian brochure there are only three explicit criteria.

1) Professionals that are working in areas related to the conservation of species of flora and fauna and their habitat.

2) Undergraduate or graduate students of Colombian universities that carry out a thesis project required for their degree and who wish to develop their thesis in an area related to conservation of biodiversity in the country.

3) "The program also will finance projects without a field component that offer some kind of help to the conservation of biological diversity."

The brochure does not mention other criteria for selection. It does not speak of the priority of quality, of the justification of the project's importance to conservation (although the student must state the importance), the budget, the location, or any other criterion. It does not say who the Scientific Committee is. Thus there is no rationalization of why one criterion may be used over another. This seeming arbitrariness can hurt program credibility. Also it says that professionals in any field can apply and then later requires that professionals deliver a letter of support from their institution. Thus are freelance or unemployed graduates excluded? This would be unfortunate since young professionals, those "freelance conservationists" (in the words of the program coordinator), may consider themselves excluded.

The third criterion really indicates that the program's criteria may be arbitrary by granting money to anyone who does anything related to the conservation of biodiversity. Does "anyone" mean students or can government employees also solicit funds? Clearly non-research programs for established professionals is not part of the defined target population.

This kind of writing would alert evaluators to observe and document the selection process. But what other factors might the Committee use in selection to make the program more finely correspond to the objectives?

Certainly the programs are intended to overcome a hopelessness that there is no money for programs. But this program is based on self-selection and thus only motivated students apply, leaving the most hopeless to remain hopeless and penniless. If the program were targeted at thesis advisors rather than students, they might be tapped to seek out those they feel are less motivated, less aggressive, or less experienced in fund raising. Also the program is designed to offer money where other sources do not exist. Presumably then financial need should be a criterion, however, the program does not ask for any financial information. The program coordinator says, moreover, that most students get loans from family or work second jobs to pay for their research and the program affords only "incremental value." What does this mean? If students do not really need the money, then why have a grants program? If they do need money, then how much? Perhaps the program is overgranting. Perhaps many financially needy are excluded to pay for richer more motivated students who do not need the money. These are empirical questions discussed later on.

Although few data exist, we know that in several selection rounds for the small grants program in Colombia there has been a pre-dominance of applicants from the Andean region with a complete absence or continual low representation of applicants from other regions such as the Pacific or Atlantic. Most requests for information come from the major urban areas surrounding Bogotá and the Cauca Valley Department. Is this desirable? Has program administration done anything to even out the representation? Since the data are poorly presented in the internal document, it is hard to draw any conclusion. An evaluator would have to re-analyze the data to the extent possible.

Presumably academic quality consisting of the idea and the methodology are of

importance, although the applicant does not know. They do not know if the budget affects the decision. I do not mean to say that these criteria should matter, perhaps the program looks to promote research that does not require much equipment because that is how it is in the field. Also the program may decide it wants to recruit or select more people from geographically distant locations or smaller universities. Or maybe it would like to get more applicants in certain fields of study because the country lacks specialists in those areas. The program has the power to fill in such gaps. Ideally these preferences would manifest in recruitment rather than selection in order to protect a reputation of fairness, although it could select using quotas or special positions designed solely for one type of applicant.

The program could emphasize greater gender equity or more college students than young professionals. Recall that the regional meeting suggested that undergraduates were less risky than professionals to train because the former have not yet made their career choice. Again empirically we can test to see if the effect on young professionals is sufficient to justify not awarding more money to undergraduates. Last, what constrains young professionals from entering the conservation field: money, experience, contacts? If the constraint is interest or motivation, then the program is missing the correct target because of self-selection again.

Looking for Bias

Coverage bias constitutes a major concern for all social programs. Quite often motivated individuals are not the targets. Some programs cater to the most qualified of their population to make their outcome statistics like good. Very often this is done accidentally. In the case of the conservation course, most of the bibliography is in English, thus there is a coverage bias against anyone who cannot read English or read it well. The program director in Colombia noted in a quarterly report that the English barrier has resulted in a lack of good candidates. Thus what has the program attempted to do to rectify this problem? Translation? Shift the bibliography to Spanish? Offer courses in reading English science literature? What groups are most disadvantaged by a lack of English in Colombia? Most assuredly poorer applicants, applicants who come from areas farther away from urban areas, those who come from lower quality universities, maybe even females more than males. In short, the course becomes elitist and biased against an important segment of its target population. An evaluator may attempt to document how applicants are turned away or how they drop out due to this barrier.

What is the interface between potential applicants and the program? How are conservation science course applicants judged? Since it is this program that has trained both high school and university professors, and no application criteria were found in the literature, an evaluator should focus on selection as a possible problem area. One question to keep in mind: do all potential target units have an equal chance of participating? In theory, they should. But it is clear already they do not.

Continuing the Implementation Evaluation

Many questions asked in this report are empirical and can be obtained through participant surveys as will be discussed later on in this report. Light can be shed by talking with key informants and comparing recruitment methodologies for other similar programs such as the grants program for Alexander von Humboldt Institute in Colombia or may some other non-environmental grants program of FES in other sectors such as health or human rights.

An evaluator must go through all program records and extract information about selection patterns, budget allocations, and exactly what data have been recorded. He might

also do a quality check by randomly selecting records to see if they were filled out accurately and completely as well as check to see if all applicants from a registration list have files. Have records of rejected applicants and those who dropped out been maintained? Are all final reports on file?

A very important system to evaluate is how alumni have been tracked and where their records are maintained. The program coordinator says the in-country administrators keep good track of most of the alumni: is this a formal effort or do they just call various alumni when they have time? Do universities track alumni? Is there duplication of records or could efforts be shared?

Certainly a semi-structured interview with university officials and other stakeholders would also be in order for many of the questions herein asked.

Last, for the conservation science course clearly an in-class evaluation would be part of an evaluation. This report will not delve into the complex matter of evaluating pedagogy, which certainly can go beyond the simple appraisal of an exit evaluation and involve pre- and post-tests, in-class observations, faculty interviews, an evaluation of the exit surveys themselves, and a comparison of course objectives with outcomes.

Fleshing Out Outcomes

Thus according to the WCS program coordinator and literature, very little data has been systematically collected. In fact the implementation has not been designed to evaluate results. More contradictions and am-biguities will likely arise as an evaluation continues. Truly meaningful outcomes cannot be measured until control groups and proper data collec-tion are initiated. Table 1 offers the list of evaluable areas for the programs from which WCS may choose to investigate.

Nonetheless the following suggestion discusses how WCS might go about investigating outcomes both in the short-term and long-term. The multi-pronged effort that in effect says, "We're trying our best now and we have a plan to do much better in the future" is a viable argument for donors.

Table 1 Goals, Programs, Time Frame

Objective/Program/Time Frame	Grants	Course	Short-term	Long-term
Provide Needed Funding	X		-funding assessment -alumni survey	-funding assessment -control group -financial aid applications -longitudinal survey
Overcome Syndrome of Learned Hopelessness/infuse optimism/achieve open mind, etc.	?	X	-alumni survey (insufficiently)	-pre/post test -longitudinal survey
More people to enter field	X	X	-control groups?	-control group
More people to enter field faster	X	X	-control groups?	-control group
More people to create own opportunities	X	X	-control groups?	-control group
Develop conservation leaders	X	X	depends on definition?	depends on definition?
Motivate interest in conservation careers	X	X	-alumni survey	-pre/post tests -compare behavior w/ post test; -longitudinal
Learn useful skills and attitudes	X	X	-alumni survey	-pre/post tests -longitudinal survey
Provide staff for national research institutes	X	X	-census institute staff	-census institute staff
Transform scientists into conservationists	X	X		-longitudinal survey
Increase networks of conservationists	X	X	-alumni survey	-longitudinal survey -census network directories
Develop local NGOs	X	X	-observation -key informant survey	-observation -key informant survey
Building professional capacity	X	X	depends on definition?	?
Driving the protection, conservation, and sustained use of the biological diversity	X	X	?	?
Provide grants	X		-tautology	
Provide guidance/ supervision to grantees	X	X	-alumni survey	

Short-Term Outcomes

When facing a lack of built-in evaluation mechanisms, short-term outcome evaluations often involve recall data from program participants, key informant interviews, and, depending on available records, some control group reconstruction.

Objectives to Measure

Certainly there are many goals that have been stated for the programs. They cannot and should not all be used. Many do not have any connection to an impact model and others conflict or are ambiguous. Table 1 outlines the major goals for the two programs. Those in italics have been operationalized to some extent. The suggested methods are by no means exhaustive; as objectives are operationalized the proper choice of methods should clarify itself.

Assessment of Funding Opportunities

A possible operationalized objective of the small grants program might be:

To provide funds to university students and young professionals who would otherwise be unable to finance their research with grants sufficient to meet at least 75% of their budgetary need.

Of course the outcome is simple enough: the delivery of the grant. Identifying need, however, is more difficult. An assessment of funding opportunities can take several forms.

1) On the alumni survey, a number of questions will get at applicants' budgetary needs and opportunities they encountered when searching for funding.
2) The evaluator can conduct a survey of university advisors asking about funding opportunities with questions such as:
    a. How many other projects were conservation-related and funded by other sources? Which sources (including personal funds)?
    b. How important, in general, is the small grants program to steering students in conservation-related careers?
    c. Are there some segments of students (less motivated, from specific departments) that the program should reach but as of yet has not?
3) A closer on-site inspection of some of the biggest universities that contribute applicants could be conducted to see where students acquire funds. Using university records, it is likely the financial sources of the previous year's thesis research can be ascertained.
4) Advisors of rejected applicants should be contacted to find out if rejectees were able to finance their projects or had to reduce their budget to be financed. Rejected applicants control for both motivation and financial need with those that were funded by the program. (It assumes rejection does not diminish their motivation.) This is perhaps the most objective measure of how the small grants programs matches up with other sources of funding.

A post-evaluation could ask what sources of funding they ended up using. This would be done immediately after the research as recall would not be reliable after very long for students who used multiple sources (only the answers from the previous two or three years on the alumni survey would be considered for this reason). Also after the research was conducted, there would be little incentive to hide the truth, unless it seriously contradicted

what they reported at some prior point. The application for selection might also ask what other sources of funding students are pursuing. It could be stated that the presence of matching funds makes funding more preferable for the small grants program. This strategy is used by the Tropical Resources Institute at Yale University who has a student research small grants program very similar to that of WCS's.

Alumni Survey

The survey will address all objectives that apply and will be given to a sample of the entire alumni body for each country. See section V for a suggested methodology and sample survey for the two programs.

Simple Field Observation

One objective though not operationalized can be measured convincingly nonetheless. Through a simple review, the goal about strengthening NGO conservation organizations can be critiqued. It is interesting to note that programs failed in both Peru and Ecuador due to "lack of organizational capability" and from being "inept" as the program coordinator put it. Also instead of strengthening NGOs, in Venezuela, WCS set up a new one, EcoNatura. Setting up rather than strengthening falls outside of the objective. Additionally in Colombia, the first administrating NGO was Grupo Ecológico GEA which due to problems lost the administration to Fundación FES (an already large and well established non-profit bank) and to the newly created EcoAndina Foundation. At first glance, the programs did not succeed in strengthening NGOs in either country.

Constructed Control Groups

While it is unlikely, perhaps several larger universities may have records and may track alumni sufficiently to construct control groups going several years back. If this is possible, then data on entry to conservation fields and rate of entry might be calculated. See control groups for long-term outcomes on the specifics.

Workshop

WCS would like to sponsor an alumni workshop which would promote networking, interaction, and interesting focal group discussion. For purposes of evaluation, its value would be anecdotal. Though anecdotal, its value could be enhanced if the survey data have already been tabulated and presented to participants as fodder for discussion. Alumni might also be able to help track down missing alumni whose identify will only be known after the survey has been conducted.

Long-Term Outcomes

WCS can offer a much more attractive proposal to donors by presenting a plan to generate long-term outcome data that directly relate to specific, well written and operationalized objectives. This report will work with the heretofore operationalized objectives with general suggestions where applicable. Long term would be any outcomes that cannot be measured

within roughly one year from when the evaluator goes to work.

Exit Evaluations

Since there is no guarantee that the program will coincide with the short-term outcomes evaluation in sufficient time to yield enough exit evaluations to analyze, this technique will be considered long-term (although short- to medium-term is more appropriate).

The four programs (2 programs x 2 countries) have existed for as long as eight years. Exit evaluations have been inconsistently kept and most likely exist only for more recent years. There is no indication that exit evaluations were conducted for the small grants programs.

Aside from regular administration of the evaluations, the evaluations themselves can be made more useful. They do not need to ask demographic data as they do now since that information should be in the applications. In addition to general questions that almost invariably turn out positive (For example, for an evaluation of the conservation science course in 1991 in Colombia there were 25 response categories using a scale from 1 to 5, 1 being very low and 5 excellent. The average response was 4.3.), questions can be tailored to investigate more specifically certain sections, uses of materials, certain talks, anywhere the instructors or observes may think such doubt could exist. In addition, the evaluations could be coupled with post-tests or be kept separate depending on size and format. But since no pre-tests exist, post-tests will not serve to evaluate short-term outcomes.

Pre/Post Tests

Pre and post tests can measure attitudes, beliefs, knowledge, behaviors, and intentions and how they change after the intervention both immediately and in the future such as a year after the program. Also post-test questions can be included on longitudinal surveys and be asked repeatedly into the future (or at least until alumni are defined as established professionals).

•Conservation Science Course
-Expectations for conservation course
-Reported behavior for pre-test to be followed up later
-Attitudes and knowledge, especially as they pertain to the SLH
-Intentions of pursuing a conservation career

In large measure, these tests will be designed in conjunction with faculty and in-country program staff.

•Small Grants
-Attitudes and intentions of entering the conservation field before and after.

The pre-test could be part of the initial application process and the post-test while short, is very useful in measuring intentions perhaps six months to a year after the termination of the project to see if the project has affected the student's thinking about the conservation field. This time series data will help to corroborate the impact model that the program intervention's

timing is crucial to its success in influencing career decision-making.

Control Groups

There are at least three different kinds of control groups that can be constructed.

a. Rejected applicants

Rejected applicants from the small grants program controls for both motivation and financial need as they come from the same pool as those granted. Quality of work is not controlled; however, it is unlikely one could find a significant difference between them based on quality as that would require the Scientific Committee to be highly objective and consistent in its selection process.

By communicating with their advisors (whose names and phone numbers are required by the applications), it should be easy to learn if their proposals were funded elsewhere or if the small grants program is fulfilling a unique need. Also by contacting them via telephone, rejectees may be willing to participate in a short longitudinal survey.

b. University Control Groups

Other non-applicant thesis researchers from a number of universities can be combined in annual cohorts to serve as controls for small grant program recipients to test how often and how quickly they enter the conservation field.

A similar control group can also be constructed for the conservation science course, not from thesis researchers necessarily but from a demographically related sub-group. Exactly what this group would look like depends on the characteristics of the applicants.

There is no reason to believe, however, that these people would not serve in a longitudinal survey. If universities track their alumni, the task is easier.

c. Young Professional Control Groups

It is uncertain how to construct a control group for young professionals since no pool of such people is readily identified. Perhaps CONICIT or other federal funding agencies have programs for young scientists that could be used to compare the treatments.

In all cases, program record keeping must be upgraded to consistently record a variety of data from participants, input them in a database, and track alumni over time. A longitudinal survey could be sent out every two years to record how alumni and controls have changed jobs and attitudes. Through regression analysis, evaluators could control for a variety of demographic traits.

Credits: Thanks to Todd Strauss at the Yale School of Management and the book, *Evaluation: A Systematic Approach* by Peter Rossi and Howard Freeman (5th edition, 1993, Sage Publications: New York).

Box 3 General Implementation Questions

Describing the Program
Does the program have a written program description of how the program should be implemented? How it is implemented?

Treatments
What are the treatments? How are the treatments applied (in full or partially)? Are the treatments applied consistently across participants?

Staff Qualifications
Who are the staff members, what are they trained to do, and what do they do? Also which staff have which responsibilities? Are staff competencies adequately utilized (under/over) by the program? Who manages the budget?

Promotion Mechanisms
Does the recruitment procedure reach the entire target population evenly? Are applicants dissuaded from participating due to the course's distance?

Optimizing Access
Does housing become an prohibitive expense? Are there other barriers to program access for certain groups of the target population?

Target Retention
What is the attrition rate of the programs? Why and who drops out of the programs? What efforts do the programs take to reduce attrition and how well does the program keep up with graduates after the termination of their program? How is program completion defined?

Program Activities
What are the activities of the staff? Of the program? Where do they take place? How are they carried out? What equipment is used? How much does it cost? What is the annual cycle? How are applicants selected?

Making the Parts Conform

Problem Definition
What is the problem? Can it be defined specifically and to the agreement of the stakeholders? Does the problem properly consider its causes and effects?
What is the nature and distribution of the problem?

Goals and Objectives
Do goals refer back to the main, defined problem? Does program literature affirm the same problem and goals or refer to diverse goals and problems?

Impact Models
Has the program explicitly identified its impact model or implicitly assumes one? Does the model made sense or have logical gaps? Is the model based on prior research or empirical data or just someone's best hunch? Is it practical to conduct such a search? What are the assumptions to the unspoken underlying model? Can the impact model be defined in terms of the three kinds of hypotheses (causal, intervention, and action)? Can a schematic of the model's operation be drawn and if so can gaps in the reasoning be identified?

Operationalizing Objectives
Does each objective contain only one aim and one outcome? If an objective contains two results and one has not been met, to what extent can the program claim to have met the objective? Are they written as clearly as possible? Does the program have a time component in it?

Covering Coverage

Defining the Target Population
Does the impact model justify the chosen target population?

Too Much, Too Little
Does the program over/undercover its target population?

Looking for Bias
Is self-selection bias causing some groups to be overcovered while others to be undercovered? Has the program noticed or done anything to even out such a bias? Are their barriers that prevent coverage to certain groups, while promoting coverage of others? What is the interface between potential applicants and the program? do all potential target units have an equal chance of participating?

Continuing the Implementation Evaluation
On record keeping: Have records of rejected applicants and those who dropped out been maintained?

Appendix: Short-term Alumni Survey

The implementation evaluation of the two programs which examined program literature has identified a multitude of potential goals. The goals have been articulated by many voices in many media over the eight years of the oldest of the four programs (2 countries x 2 programs). Thus no specific goals can direct the form and purpose of this survey. Nonetheless, I have identified several important, and perhaps more measurable goals, that serve as the pillars supporting the structure of this survey. They include the following for each program (assumes that programs are implemented identically in both countries):

Conservation Science Course

1) To get more people to enter the conservation field (2)
2) To get people to create conservation jobs that did not exist before (3)
3) Overcome Syndrome of Learned Helplessness (8, 9, 10)
4) Learn certain skills, attitudes, and knowledge (11)
5) Increase intention of entering the conservation field (5)
6) Facilitate networking (4)

Small Grants Program

1) To get more people to enter the conservation field (2)
2) To get people to create conservation jobs that did not exist before (3)
3) Provide funding where none other is available (7, 9, 10, 15)
4) Increase intention of entering the conservation field (5)
5) Facilitate networking (4)

Various implementation questions: (Both: 1 [promotion], 6 [coverage]; small grants: 11 [consistency of treatment application], 12 [image building], 13 [targeting population, coverage]; course: 13 [coverage]; 14–18 [targeting population])

Applicable survey questions in parentheses.

Population and Data Collection

There are four independent populations to be surveyed. Supposedly in-country non-governmental organizations (NGOs) have at least the following information for all elements of the population: name; city and province of origin; sex; program from which they graduated; when graduated; if they were university undergraduates, graduates, or newly graduated professionals; affiliated institution; and who dropped out before completion of the program. For the small grants program we know what their project was, whether they fulfilled their obligation of delivering a final report, and how much money they were awarded. Because this information will be analyzed with survey data, the survey will not be anonymous. This should not deter many, if any, alumni because the questions are not very personal and administrators and alumni have good familiarity.

Since alumni are scattered about the country and some are studying abroad and since mail is unreliable in more remote areas, a telephone interview will be employed. As the program is prepared to invest the resources necessary to track down missing alumni, we can expect to locate almost the entire population. Also because the population is small and has a vested interest in this program, we expect a very high rate of response, in the 90s%. Thus we will attempt to reach all alumni through repeated callbacks and aggressive

investigation. Because the population is so small and whatever systematic bias that might prevent alumni from participating would be prevalent whether or not a sample of the population is used, we will take advantage of all available elements by trying to interview all alumni.

Basic Survey design

With the assumption of identical implementation (confirmable or deniable only upon investigation in-country), the two surveys will be combined in the same document. The first section has questions applicable to both programs; the second and third sections refer to the small grants and conservation science courses respectively; the fourth section contains demographic questions for both programs. The entire survey should last 15–20 minutes.

While the survey was pre-tested on non-target people, because the target population cannot be completely simulated in pre-testing, the survey will also require pre-testing with real alumni. Due to the small population size only as few alumni as necessary will be used for pre-testing, starting with a minimum of five and more will be interviewed depending on the rate of changes being made. Data will be kept for questions which are not modified.

Analysis

- Data on entering the field will be graphed "years out of program" versus "number of alumni." Curves for both programs and control groups will eventually be graphed together. These data are baseline. We can also graph how soon jobs (first and second) were created and what frequency by program and number of alumni.
- Recall will be tested for its validity at least in terms of financial data. Only need to time alumni to the point where we can identify the relationship between the rating score that alumni use and the operationalization that we have here.
- Aside from summary statistics on collaboration, we can begin to assemble a collection of collaborative anecdotes useful for program promotion.
- Intentions can be measured against what actually happened to indicate the degree to which intervening factors are limiting the realization of the intention. In the future when intentions are not recall but actual, this information will be more meaningful.
- Using real percentages and estimated percentages of funding, we will try to understand how important the grant really was in funding. Again recall likely reduces the quality of these data. In the future they will be collected at the time of the grant.
- The kinds of advice, guidance, and suggestions received will be post-coded.
- In the future pre/post program data will be analyzing looking for significant differences. For now we must rely on recall data until those reflexive data exist.