A Program Evaluation Primer

Simon Priest

Introduction

With a few exceptions, programs from the experiential learning discipline and its associated fields of adventure programming and outdoor education have been poorly evaluated. This simple primer explains the five models of program evaluation and how they differ from research. Insufficient space precludes providing step-by-step instructions on how to conduct evaluations in this article, but enough information is presented to get practitioners started with their evaluations.

The initial and most important concept to understand is that evaluation and research are very different. Most practitioners are scared away from evaluation because they see it as a form of research that requires the expertise of an academic "rocket scientist"! In truth, evaluation is more basic and less time-consuming than research. Whereas research attempts to "prove" that a practice is effective, evaluation seeks to "improve" the effectiveness of that practice (Borg & Gall, 1983). Evaluation can be as simple as counting how many people participated in a program and for how long. Evaluation can be moderately difficult like surveying satisfaction levels with the program. Evaluation can be as complex as observing how participants have changed after a program.

Definitions of Key Terms

A **program** is a collection of several learning experiences held together by logistics such as scheduling, staffing, equipment, meals, housing, transportation, communication, finances, and so on. For the purpose of evaluation, think of the program as including these elements and all others that support the learning experiences and not just the experiences alone. Examples include an indoor team-building retreat for corpora-

tions, an outdoor adventure day camp for children, a wilderness expedition for survivors of abuse, and a course in rock climbing for the general public.

A **learning experience** is a specific and individual event in which people engage (directly or indirectly resulting from the program) that subsequently changes the way they feel, think, or behave. Examples matching those programs listed above include a problem-solving task or unexpected conflict among group members in the team building, a canoe trip or getting lost on a hike at the day camp, fording a river or pitching a tent in the dark during the expedition, and tying knots or falling on belay while rock climbing. Some of these are planned parts of the program and others arise spontaneously from the circumstances.

Experiential involves the addition of the following elements to the base experience: reflection, integration, or continuation. After participating in the experience, people reflect on it to learn particular lessons. If they transfer those lessons from the experience to their daily lives, then their changes are likely to be integrated. If their changes in feeling, thinking, or behaving are supported in daily living, then their learning is continued.

Research is a type of inquiry that uses systematic methods and rigorous procedures derived from scientific knowledge and logical thinking to "prove" whether a practice is effective in generalized situations. For example, a study that examines how trust is consistently developed in most groups across many programs would be a form of research.

Evaluation is another type of inquiry that uses similar research-related skills to "improve" practice effectiveness in a specific situation. For example, a study that examines whether trust actually develops in a single group in one program would be a form of evaluation.

Generalizability is the degree to which research or evaluation findings from a study can be applied to other situations that were not studied. Generalizability is the principal difference between these two types of inquiry. Research seeks a higher amount of generalizability than evaluation and so has stringent requirements about selecting or controlling subjects, treatments, variables, and measurements within a study. Again, fear of

Simon Priest, Ph.D., is CEO of virtualteamworks.com: a collaborative collective that builds virtual teams, develops global leaders, and trains electronic facilitators using online communication and internet learning experiences.

research and its precise requirements for generalization unfortunately prevents many practitioners from conducting evaluations.

Experiential program evaluation is the use of research-related skills to ascertain the goals, objectives, viability, accuracy, success, utility, worth, effectiveness, quality, and impact of a practice, experience, or program. The purposes of experiential program evaluation are:

- to determine the program objectives and why these are important;
- to decide (beforehand) if a program can meet these objectives;
- to monitor and recommend (midstream) adjustments to the program;
- to examine (afterwards) whether the program was indeed effective;
- to calculate the program's value in relation to its costs and benefits;
- to investigate how these elements and practices can be improved; and
- to enhance the evaluator's experience of conducting evaluations.

Groups of Stakeholders

Six principal groups hold a stake in contributing to and being rewarded by the findings of evaluation. These stakeholders are classified by whether they provide or consume the program:

- **providers** include suppliers, staff, and several specialist professions; and
- **consumers** include customers, clients, and their greater communities.

Suppliers organize the program and, thus, influence its format. Repeat business depends upon the quality of their programs. **Staff** facilitate the program and, therefore, impact its content. They have a vested interest in offering high-quality programs to maintain their jobs. **Professions** govern the standards, operating procedures, or ethics of programs and providers. These professions stand to lose or gain credibility on the basis of the program quality offered by providers under their influence.

Customers pay for a program and expect the quality to be commensurate with the provider's advertising. The customers benefit from fulfilling an obligation such as training employees, rehabilitating addicts, or educating children. **Clients** participate in the program and their degree of involvement and motivation can ultimately determine the quality and success of a program. They gain directly from learning and change derived from the program. **Communities** (such as a corporation, judicial

system, or society) create the customer's obligation to purchase a program. These communities prosper indirectly from changes such as increased profits from productive employees, reduced crime and violence due to diminished addictions, or improved lifestyles resulting from more creative and inventive citizens.

Unlike research, which might focus on one of these groups, evaluation should consider and include all stakeholders. For example, providers (suppliers, staff, and professions) shape the nature of programs. Similarly, consumers (customers, clients, and communities) give input to ensure that programs meet their specified needs. Consequently, satisfied consumers enhance the long-term viability of providers by bringing new business and promoting programs to other markets. Lastly, sharing evaluations and effective programming ideas among all stakeholders leads to better provider practice and more successful programs for consumers.

One additional holder of a stake in the evaluation is the actual **evaluator**. In research, the researchers must be the most primary stakeholder in order to maintain control over the program studied. However, in objective evaluations, evaluators should be last to influence programs.

Reasons for Doing Evaluations

Demands of the program determine the reasons and methods for conducting evaluations. These reasons can be categorized into one of three classes (Isaac & Michael, 1983):

- accountability (to confirm that objectives are met, to make better decisions about program planning or operations, to authorize fiscal payments, to meet grant obligations, and/or to correctly allocate program resources);
- improvement (to identify program strengths or weaknesses, to create safer practices, to increase educational value, to enhance competence, to test innovative novel ideas, to diminish planning problems, to decrease operating costs, to reduce staff concerns, and/or to establish quality benchmarks or assurance standards); and
- marketing (to advertise past program effectiveness, to indicate a collective track record of successful programming, to promote positive public relations, and/or to advocate or lobby social policy).

Accountability is driven by internal and external demands. An outside funding source frequently wants some evidence that money is being well spent. Inside administrators want affirmation that staff and resources are being employed properly. Both these stakeholders want to know if the program is effective, where it might be improved, and whether learning and changing are

really taking place. Evaluation provides these answers.

Improvement is the internal demand of making practice better by identifying program shortcomings and taking action to correct these. The results of this evaluation may lead to training staff, upgrading equipment, changing activities, adjusting schedules, or even switching programs or customers.

Marketing is the external demand of demonstrating that programs can work. This is an important ethical distinction that practitioners must make. Although research might easily show that programs are effective and that they may well be similarly effective in other generalized situations, evaluation cannot make these same claims. All evaluation can suggest is that the program has a history of being effective in specific past settings. The prediction or stretch to future programs must be left to the consumer.

Methods for Conducting Evaluations

Methodologies are the procedures used to collect, analyze, and report sets of data (important information) that will aid in completing the study and answering the initial evaluation questions.

Methods are classified into one of two categories (Lincoln & Guba, 1985):

- quantitative (statistical analysis of numbers and amounts); and
- qualitative (description of distinguishing patterns or characteristics).

In **quantitative** methods, numbers typically come from an evaluator's measurement of the stakeholders' responses to surveys or standardized tests and/or from an evaluator's notes about counts of the stakeholders' actions. These measured numbers (data) are subjected to statistical analysis by the evaluator in order to describe what this information "looks like." The evaluator accurately reports these statistical outcomes in descriptive form, but without generalizing to situations that were not studied. Evaluation statistics are solely descriptive and only include the simple procedures of frequency (percentages, histograms, or distributions); central tendency (mean, median, or mode); and variability (range, variance, or standard deviation). None of the inferential statistics from research (such as t-tests, analyses of variance, correlations, regressions, or crosstabulations) are necessary in evaluation, as these procedures are more useful to ascribe findings from the research study to other generalized situations or from samples to populations, and to imply significant change or relationships. Evaluation does not include generalizing, sampling, or establishing significance, and so these inferential procedures are not necessary.

In **qualitative** methods, information is collected in detailed field notes taken by an evaluator interviewing and observing the stakeholders' actions and/or reading their written comments on surveys or in personal journals. These notes and writings are compared and contrasted for common content by the evaluator looking for patterns of similarity, difference, and relationships in the data. The evaluator interprets the meaning of these patterns in the context of the stakeholders' cultures and the program studied. The evaluator reports these interpretations in rich and thick description by using personal and subjective language (Patton, 1990).

Models of Evaluation

The five models of program evaluation and the primary questions they address are:

- Needs Assessment: What are some gaps that the program will fill?
- Feasibility Study: Given the constraints, can the program succeed?
- Process Evaluation: How is the implemented program progressing?
- Outcome Evaluation: Were program goals and objectives achieved?
- Cost Analysis: Was the program financially worthwhile or valuable?

Typically, the first two happen well before a program is delivered, the third occurs during the program, and the last two are completed after the program has finished (Priest & Klint, in press). For example, a residential center for young offenders is offering an outdoor tripping program as an alternative to prison. A **needs assessment** is conducted with the youth and other stakeholders to identify the gap (objectives) between where they are (existing situation) and where they would like to be (desired potential situation). On the basis of these needs, a program is planned to fill the gap between their current and changed positions. A **feasibility study** is conducted to determine the likelihood of success for this program by considering what is probable and what is not possible due to legal restrictions, funding shortcomings, and time, staff, or resource limits. As the program is implemented, **process evaluation** is used to track its progress by examining how and when the program is adjusted to suit changing stakeholders' needs and to accommodate staff flexibility. After the program is finished, **outcome evaluation** is employed to decide whether the stakeholders are satisfied with the changes; and once released from the center, the youths' rate of offending is monitored. Lastly, cost analysis correlates the price and benefits of outdoor tripping with the expense and rehabilitative worth of prison life. One other related form of inquiry is:

- Research (generalizability): Will the program work elsewhere?

Although research has a large number of other applications, its primary association with evaluation is to "spin off" programs to other situations. For example, if the program noted above were considered to be effective, then **research** might be utilized to decide whether the program's success could be replicated in other locations or extended to other age groups.

The Programming Sequence

In order to be effective, evaluation cannot be an afterthought; it must be a planned part of all five interconnected stages of a programming sequence (Priest, Gass, & Gillis, 2000):

- diagnosis (performed in advance to establish program objectives);
- diagnosed objectives);
- delivery (executing the program design with flexible adjustments);
- debriefing (adding reflection and integration of lessons learned); and
- disembarkation (including integration and continuation of change).

Diagnosing refers to assessing the needs of clients, customers and the community. Two simple rules of thumb for conducting diagnoses are to seek multiple sources of information and to use multiple methods to gather that information. In other words, don't ever rely on one person's opinion about what needs to be done and be sure to utilize a combination of survey, observation, and interview methodologies.

Designing is simply planning the program based on what was learned from the earlier diagnosis. In addition to considering logistical concerns, this stage determines the program's purpose and the sequence of learning experiences. The design stage also sets boundaries for what will be permissible: customers are contracted for desired breadth of change and clients are negotiated with for allowable depth of learning. Ground rules and operating principles are established and staff roles are clarified early.

Delivering is the presentation of the designed learning experiences in a program sequence. This often involves a number of critical facilitation methods for staff such as frontloading, intervening, framing, managing information, observing behaviors, and remaining flexible to change the program in midstream and as necessary.

Debriefing is the most common form of facilitated reflection, where clients look back on the delivered learning experiences and extract key lessons that they can integrate and transfer to their daily lives. This usually takes the form of a discussion where staff ask several carefully ordered questions using a technique known as funneling, and sometimes this may be coupled with a solution-focused form of questioning.

Disembarking refers to the final stage where clients leave the program and have various reactions to final closure and the continuation of their newly integrated learning. This stage involves action planning, engaging metaphors, anchoring the experiences, and facilitator feedback.
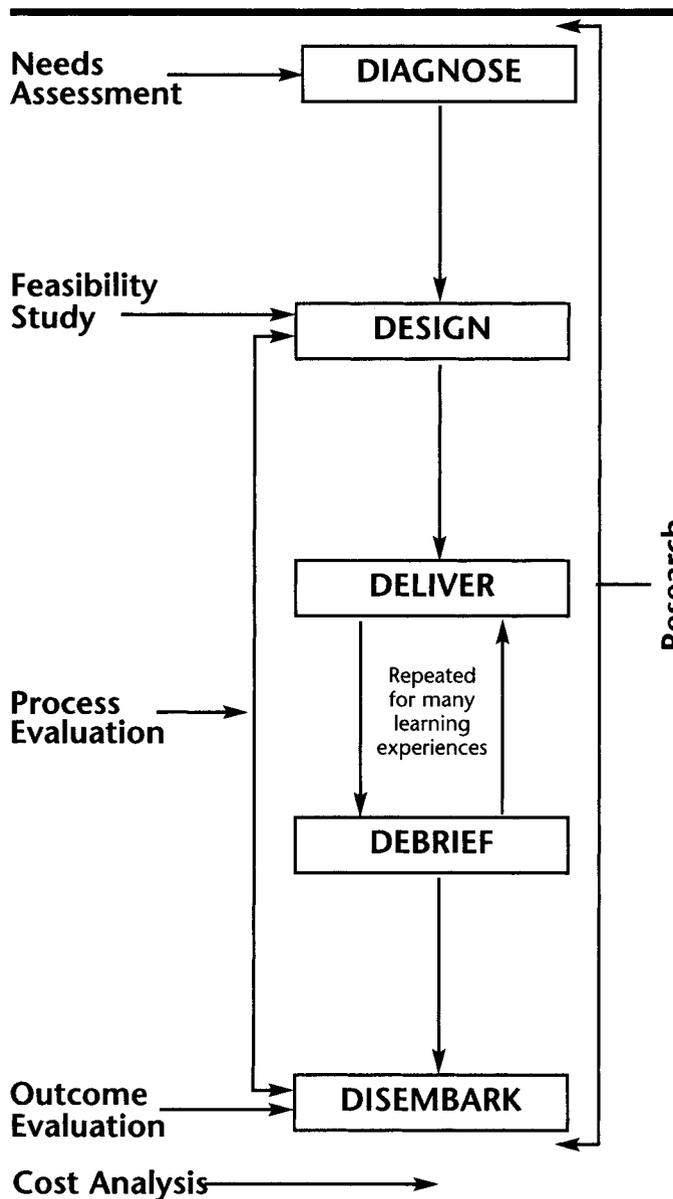
Each of the five kinds of program evaluation fits into the programming sequence of five stages as positioned in Figure 1. Notice that research can and does occur in all five stages. Table 1 compares and contrasts the five kinds of program evaluation on the characteristics of when and where each fits in the programming sequence, what each measures, how and why each is conducted, and who contributes to and benefits from each.

Needs Assessment

Needs assessment measures the gap between what is and what should be. "What is" refers to the present state of affairs and "what should be" refers to the desired target state that consumers would like to reach. The results are used to understand the context of the program (the current culture, climate, and concerns of clients, customers, and their community) and to set program goals and learning objectives. These goals and objectives inform and direct the program planning process in the design phase. For example, a corporate customer comes to a program provider with a general request for a team-building session, but cannot provide any specifics on what should be addressed during the program. The needs assessment determines the client's current levels of teamwork and identifies areas where they and others see room for improvement.

Identifying the present state of affairs requires that the context of the clients, customers, and their community be fully understood by surveying, interviewing, and observing these people. Identifying the desired target state requires that these people disclose how they would like to see things change as a result of the program. After the gap between these two states has been measured, the program can be planned to bridge that gap and change the present state of affairs into the desired target state. An important piece of identifying states and measuring the gap between them is being able to write goals and objectives, recognize opportunities, tease out underlying problems or concerns, and indicate unrealized needs. For example, the corporate customer might request a team-building program that deals with miscommunications, but on closer examination, the evaluator finds that the clients are miscommunicating

Figure 1: Sequenced positions of the five kinds of program evaluation.

the factors that might help or hinder the program delivery. Special attention is often paid to the best use of resources (such as staff, equipment, finances, space, time, etc.). Everything is considered within the social climate and present or expected circumstances that exist and may influence these resources. For example, an adventure therapy program may be proposed, but funding sources may prefer a more punitive style of "boot-camp" instead. These two should be compared and contrasted in an effort to determine which is most feasible.

Feasibility studies identify alternative approaches and helping or hindering factors. All realistic alternatives are compared with the program design to determine whether the latter is indeed the best way to go. Helping and hindering factors are inventoried to decide whether the program is likely to work given the barriers and resources that exist. Not only do these factors account for supports (such as staff, equipment, finances, space, time, etc.), but also social, political, legal, moral, and fiscal restrictions. For example, one can design an amazing adventure therapy program but may lack the wherewithal to make it work, or may face societal prejudices that will block its ultimate implementation in favor of the military model. A more conventional "boot-camp" approach may be more readily accepted or may be better suited to address the customer's demand than the adventure therapy program that might better address the client's needs.

Process Evaluations

Process evaluations measure the gap between the program plan and its actual execution. The program is monitored to ascertain whether its delivery matches its design and to determine if midstream corrections are called for. This information is used to modify and improve the program as necessary during its delivery. For example, during a wilderness expedition, many things can derail the trip. A washed-out river bridge or a lengthy argument among clients can lead to a change in plans and the need to drop some parts of the program or add others.

Process evaluations contrast the needs from the earlier needs assessment with all new needs that arise within the program. They compare the content and format of what is happening with what was planned. They result in adjustments that refine the content and format to better address the new needs that have arisen. They require that staff remain flexible and are quick to deviate from the plan when doing so is in the client's best interest. For example, clients on a wilderness expedition are interested in feeling better about themselves, but the staff discover that this cannot happen without their venting of past negativity. Unfortunately the program does not allow for this opportunity. Nevertheless, staff

because they don't trust one another. Not only does the needs assessment highlight the gaps in communication and trust, but it also notes the relationship between these two elements of teamwork and other pieces such as cooperation or collaboration.

Feasibility Studies

Feasibility studies measure how likely the program is to succeed by examining alternative approaches and

Table 1: Comparison of the Five Kinds of Program Evaluation

	Needs Assessment	Feasibility Study	Process Evaluation	Outcome Evaluation	Cost Analysis
Sequence position	during diagnosis, but before design	during design, but before delivery	during delivery and/or debriefing	during and/or after disembarkation	after program completion
Measures	gap between what is and what should be	alternate approaches, help/hinder factors	gap between program plan and execution	satisfaction levels, objectives attainment	comparative merit/worth
Questions asked	what are objectives, priorities, and needs?	Which strategies and program procedures?	Are strategies and procedures working?	Are objectives met? Are people satisfied?	Should program be continued?
Get input from	clients, customers, and community	staff, supplier, clients, and customers	staff, supplier, clients, and customers	clients, customers, and community	staff, supplier, and profession
Answers used to	understand context and direct planning	gauge viability and best use of resources	monitor and modify program (midcourse)	improve/justify effectiveness	decide on future offerings
Results used by	staff and supplier	staff, supplier, and customer	staff and supplier	staff, supplier, clients, and customers	staff, supplier, and customer
Conducted by	describing context and comparing actual circumstances with intended change state	inventorying any resources or barriers and by examining all realistic possibilities	comparing arising with anticipated need and content or format with intended design	comparing actual result or product with expected outcome or standard benchmark	comparing cost ($) with benefit, effect, utility, and efficiency
Other Considerations	goals vs. objectives, unused opportunities, underlying problems, and unrealized needs	identify legal, moral, political, and fiscal restrictions, supports, constraints, or limits	identify weaknesses and strengths, remain flexible, and suggest quick adjustments	baseline measures may need to be taken if trying to measure change over time	compare with other programs, repetition ease, subjective value
Related forms of evaluation (A.K.A.)	context eval. objectives eval. demand eval. discrepancy eval.	input eval. planning eval. practicality eval. comparative eval.	formative eval. progress eval. implemented eval. transactional eval.	summative eval. impact eval. product eval. performance eval.	cost–benefit cost–effect cost–utility cost–efficiency

are able to evaluate the process and realize that the plan must be changed to allow clients the chance to "let off some steam!"

Outcome Evaluations

Outcome evaluations measure whether learning objectives were achieved and if clients, customers, and the community are satisfied with products and performances. The results of this evaluation are used to justify overall effectiveness and to suggest areas for program improvement. For example, in a public rock climbing course, clients might fill out a form that gauges their per-ception of the equipment or instruction. This might lead to newer ropes or smaller class sizes. A test of their understanding might indicate they are very knowledgeable climbers, but a test of their skill may suggest they are less competent than expected at climbing. The end result might be more practical sessions and less theoretical study in the next rock climbing course.

Outcome evaluations identify what satisfaction, learning, or change was obtained versus what was expected. They compare the actual products or performances against a standard benchmark or an earlier baseline measure. For example, the climbers' satisfac-

tions with equipment and instruction are measured against the standard of new gear and other instructors they have known throughout their lives. Their understanding and skill in rock climbing may be contrasted against their knowledge and competence as beginning climbers at the start of their program.

Cost Analyses

Cost analyses measure the value of a program in comparison with other alternative approaches (often those noted in an earlier feasibility study). Comparisons of relative merit or worth are used to decide whether the same program will run again, or be terminated or duplicated elsewhere. For example, the after-school day camp has been shown to be effective at socializing children. However, so have competitive sports and special interest clubs. A cost analysis compares and contrasts these programs in light of their common denominator: money. It determines which has the best benefit, effect, utility, and/or efficiency for the price of the program.

Cost analyses identify the price and the benefit, effect, utility, or efficiency of several programs and compare all of these using a subjective monetary denominator. For example, if children pay $20 per day to attend the day camp and the sponsoring agency pays a daily subsidy of $50 per child, then the price of the day camp is $70 per person per day. Similar calculations are performed for the competitive sports and special interest clubs. Next, the benefit, effect, utility, or efficiency of these three options are judged in terms of return on the initial investment and dollar figures are assigned to each of these. A simple profit-versus-debit comparison shows which program is the best value for the cost.

Ethics

Inadvertently misleading a few stakeholders is an unfortunate pitfall of evaluation. Deliberately misconstruing evaluation findings is **unethical** and all too common in our field. Society's appetite for information means that stakeholders will be inclined to **overgeneralize** evaluation findings because people are used to get-

ting information presented this way in the news and other forms of public media. The responsibility of an ethical evaluator is to accurately report the findings and caution stakeholders not to apply these to anything other than the situation studied.

As an evaluator, if you are asked to do a study that involves generalizing, predicting, controlling, sampling, hypothesizing, or cloning programs, then you are being asked to conduct research. If you are not qualified to do this, you have two ethical options: get experience (read books and take courses) and/or work with "experts" (consultants or supervised graduate students).

Lastly, the concern for ethical evaluation means evaluators will be expected to honor the rights of the stakeholders. Not surprisingly, these rights are very similar to those of ethical experiential programs. For example, "challenge by choice" suggests that stakeholders can select their own level of engagement with an evaluation after having had the purpose, risks, and responsibilities explained to them. They do not have to participate at all, can withdraw at any time without penalty, and can pick the questions they want to answer. In most situations, they are entitled to anonymity (no one will know who they are) and confidentiality (no one will know what they said or did). If they request it, they should receive full disclosure (written and verbal explanations of anything they don't understand) and they should be given a copy of the evaluation findings at the proper time.

A Challenge

The professional image and educational efficacy of this field is at risk because we have collectively failed to study what we do. Practice and credibility will only improve through direct evaluation by practitioners. Anyone who practices experiential learning can and should be an evaluator of its programs. Therefore, by virtue of your role as a practitioner, you are best suited to conduct the much needed evaluation of these programs. Please rise to the professional challenge!

References

Borg, W. R., & Gall, M. D. (1983). *Educational research: An introduction.* New York: MacKay.

Isaac, S., & Michael, W. B. (1983). *Handbook in research and evaluation.* San Diego, CA: EdITS.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry.* Beverly Hills, CA: Sage.

Patton, M. Q. (1990). *Qualitative evaluation methods.* Beverly Hills, CA: Sage.

Priest, S., Gass, M. A., & Gillis, H. L. (2000). *The essential elements of facilitation.* Dubuque, IA: Kendall-Hunt.

Priest, S., & Klint, K. (in press). *Experiential program evaluation.* Dubuque, IA: Kendall-Hunt.