1

EVALUATION AND TYPES OF EVALUATION

1. REASONS FOR CONDUCTING EVALUATIONS

The notion of evaluation has been around a long time—in fact, the Chinese had a large functional evaluation system in place for their civil servants as long ago as 2000 B.C. In addition to its long history, evaluation also has varied definitions and may mean different things to different people. Evaluation can be seen as synonymous with tests, descriptions, documents, or even management. Many definitions have been developed, but a comprehensive definition presented by the Joint Committee on Standards for Educational Evaluation (1994) holds that evaluation is "systematic investigation of the worth or merit of an object."

This definition centers on the goal of using evaluation for a purpose. Accordingly, evaluations should be conducted for action-related reasons, and the information provided should facilitate deciding a course of action.

Why should NSF grantees do evaluation? There are two very important answers to this question. First and foremost, evaluation provides information to help improve the project. Information on whether goals are being met and on how different aspects of a project are working are essential to a continuous improvement process. In addition, and equally important, evaluation frequently provides new insights or new information that was not anticipated. What are frequently called "unanticipated consequences" of a program are among the most useful outcomes of the assessment enterprise.

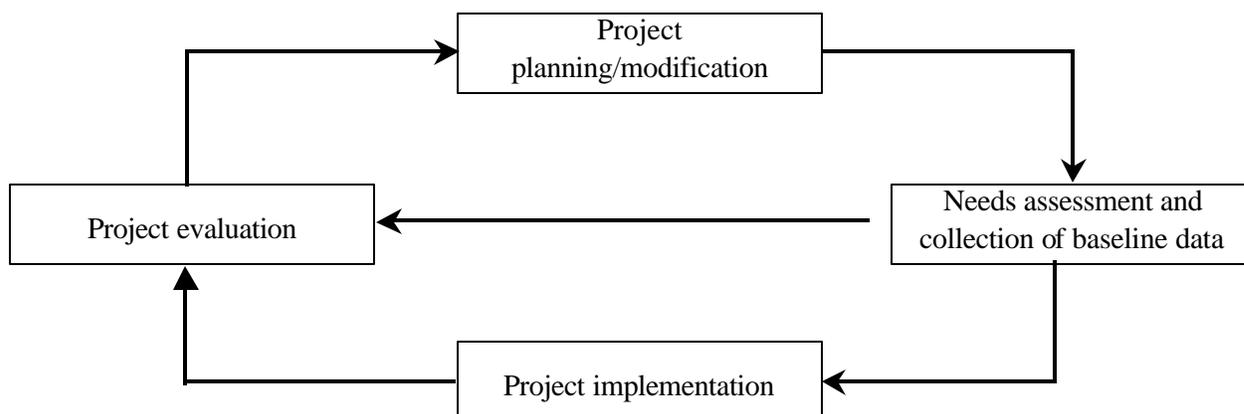*Evaluation provides information to help improve a project.*

Over the years, evaluation has frequently been viewed as an adversarial process. Its main use has been to provide a "thumbs-up" or "thumbs-down" about a program or project. In this role, it has all too often been considered by program or project directors and coordinators as an external imposition that is threatening, disruptive, and not very helpful to project staff. While that may be true in some situations, evaluations need not be, and most often are not, conducted in an adversarial mode.
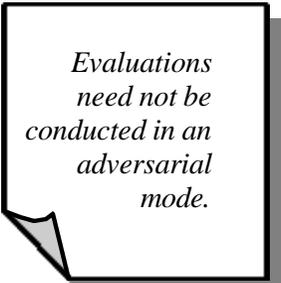
Evaluations need not be conducted in an adversarial mode.

The current view of evaluation stresses the inherent interrelationships between evaluation and program implementation. Evaluation is not separate from, or added to, a project, but rather is part of it from the beginning. Planning, evaluation, and implementation are all parts of a whole, and they work best when they work together. Exhibit 1 shows the interaction between evaluation and other aspects of your NSF project.

Exhibit 1.—The project development/evaluation cycle

```
                      ┌──────────────────────┐
          ┌──────────▶│      Project         │──────────┐
          │           │ planning/modification│          │
          │           └──────────────────────┘          │
          │                                              ▼
┌──────────────────┐                        ┌──────────────────────┐
│                  │◀───────────────────────│ Needs assessment and │
│ Project evaluation│                       │collection of baseline data│
└──────────────────┘                        └──────────────────────┘
          ▲                                              │
          │           ┌──────────────────────┐          │
          └───────────│ Project implementation│◀─────────┘
                      └──────────────────────┘
```

Evaluation provides information for communicating to a variety of stakeholders.

Second, evaluation provides information for communicating to a variety of stakeholders. It allows projects to better tell their story and prove their worth. It also gives managers the data they need to report "up the line," to inform senior decisionmakers about the outcomes of their investments. This notion of reporting on the outcomes of federal investments has received increased emphasis over the last several years with the establishment of the Government Performance and Results Act (GPRA). GPRA requires federal agencies to report annually on the accomplishments of their funded efforts. This requirement includes establishing broad goals or strategic outcomes, performance outcomes, and performance indicators against which progress will be assessed. GPRA goes beyond counts of who is funded or who is served, placing the focus instead on results or impacts of the federal investment. In response, NSF has chosen to focus on three general strategic outcomes:[1]

- Developing a diverse internationally competitive and globally engaged workforce of scientists, engineers, and well-prepared citizens;

- Enabling discoveries across the frontiers of science and engineering connected to learning, innovations, and service to society; and

- Providing broadly accessible, state-of-the-art information bases and shared research and education tools.

Projects will be asked to provide data on their accomplishments in these areas, as relevant. Detailed requirements for the information to be provided have been developed on a program-by-program basis.

---

[1] NSF, FY 2002 GPRA Performance Plan, April 19, 2001, p. 2.

Project directors should keep GPRA and these strategic outcomes in mind in developing plans for project evaluation (more information on NSF's approach to GPRA can be found at www.nsf.gov/od/gpra/start.htm).

2. EVALUATION PROTOTYPES

The purpose of this chapter is to provide a grounding in evaluation and to discuss the kinds of information evaluation can provide. We start with the assumption that the term "evaluation" describes different models or data collection strategies to gather information at different stages in the life of a project. A major goal of this chapter is to help project directors and principal investigators understand what these are and how to use them.

As we undertake this discussion, it is important to recognize that within NSF there are two basic levels of evaluation: program evaluation and project evaluation. While this handbook is directed at the latter, it is important to understand what is meant by both. Let's start by defining terms and showing how they relate.

A **program** is a coordinated approach to exploring a specific area related to NSF's mission of strengthening science, mathematics, and technology. A **project** is a particular investigative or developmental activity funded by that program. NSF initiates a program on the assumption that an agency goal (such as increasing the strength and diversity of the scientific workforce) can be attained by certain educational activities and strategies (for example, providing supports to selected groups of undergraduate students interested in science or mathematics). The Foundation then funds a series of discrete projects to explore the utility of these activities and strategies in specific situations. Thus, a program consists of a collection of projects that seek to meet a defined set of goals and objectives.

Now let's turn to the terms "program evaluation" and "project evaluation." A **program evaluation** determines the value of this collection of projects. It looks across projects, examining the utility of the activities and strategies employed. Frequently, a full-blown program evaluation may be deferred until the program is well underway, but selected data on interim progress are collected on an annual basis. **Project evaluation**, in contrast, focuses on an individual project funded under the umbrella of the program. The evaluation provides information to improve the project as it develops and progresses. Information is collected to help determine whether the project is proceeding as planned and whether it is meeting its stated program goals and project objectives according to the proposed timeline. Ideally, the evaluation design is part of the project proposal, and data collection begins soon after the project is funded. Data are examined on an ongoing basis to determine if current operations are satisfactory or if some modifications might be needed.

Project evaluations might also include examination of specific critical components, as shown in Exhibit 2. A component of a project may be a specific teacher training approach, a classroom practice, or a

governance strategy. An evaluation of a component frequently looks to see the extent to which its goals have been met (these goals are a subset of the overall project goals), and to clarify the extent to which the component contributes to the success or failure of the overall project.

Exhibit 2.—Levels of evaluation
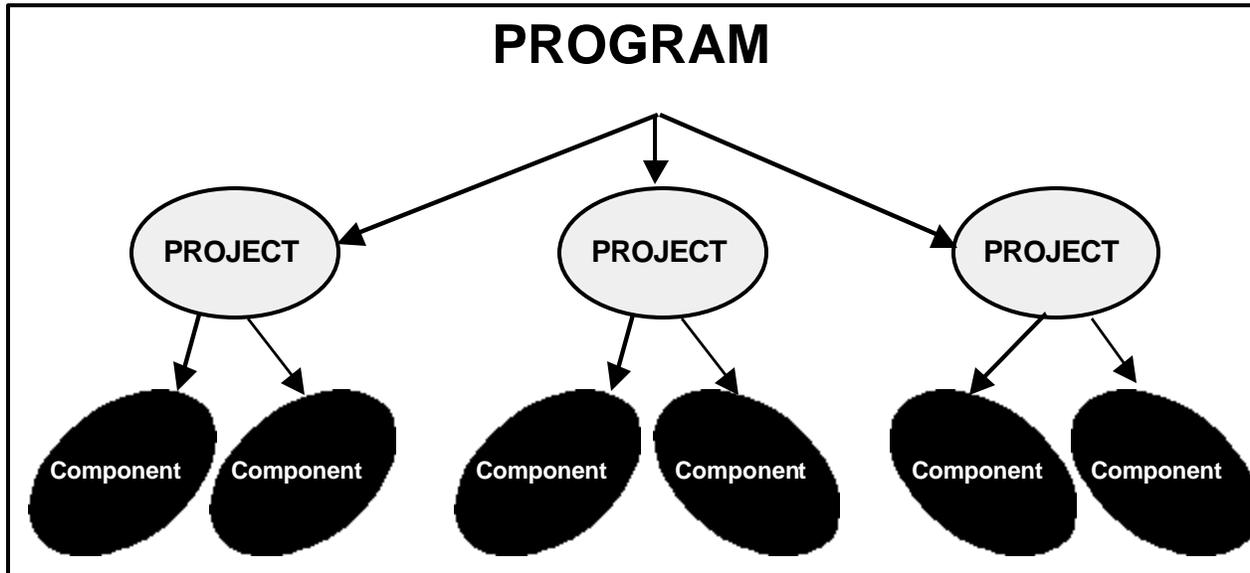
The information in this Handbook has been developed primarily for the use of project directors and principal investigators, although project evaluators may also find it useful. Our aim is to provide tools that will help those responsible for the examination of individual projects gain the most from their evaluation efforts. Clearly, however, these activities will also benefit program studies and the work of the Foundation in general. The better the information is about each of NSF's projects, the more we can all learn.

The Different Kinds of Evaluation

Educators typically talk about two kinds or stages of evaluation— formative evaluation and summative evaluation. The purpose of a formative evaluation is to assess initial and ongoing project activities. The purpose of a summative evaluation is to assess the quality and impact of a fully implemented project (see Exhibit 3).

Exhibit 3.—Types of evaluation
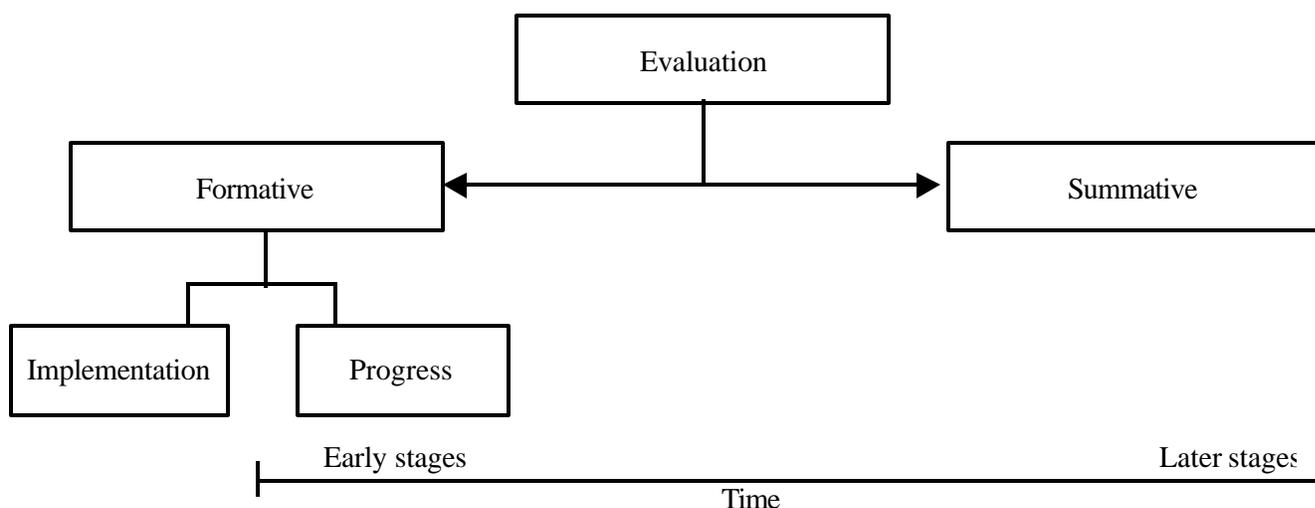
Formative Evaluation

Formative evaluation begins during project development and continues throughout the life of the project. Its intent is to assess ongoing project activities and provide information to monitor and improve the project. It is done at several points in the developmental life of a project and its activities. According to evaluation theorist Bob Stake,

A formative evaluation assesses ongoing project activities.

> "When the cook tastes the soup, that's formative;
>
> When the guests taste the soup, that's summative."

Formative evaluation has two components: implementation evaluation and progress evaluation.

Implementation Evaluation. The purpose of implementation evaluation is to assess whether the project is being conducted as planned. This type of evaluation, sometimes called "process evaluation," may occur once or several times during the life of the program. The underlying principle is that before you can evaluate the outcomes or impact of a program, you must make sure the program and its components are really operating and, if they, are operating according to the proposed plan or description.

The purpose of implementation evaluation is to assess whether the project is being conducted as planned.

A series of implementation questions guides an implementation evaluation. For example, questions that might be posed for the NSF Louis Stokes Alliances for Minority Participation (LSAMP) are as follows:

- Were appropriate students selected? Were students with deficits in precollege preparation included as well as ones with stronger records? Was the makeup of the participant group consistent with NSF's goal of developing a more diverse workforce?

- Were appropriate recruitment strategies used? Were students identified early enough in their undergraduate careers to provide the transitional supports needed?

- Do the activities and strategies match those described in the plan? Were students given both academic and personal supports? To what extent were meaningful opportunities to conduct research provided?

- Was a solid project management plan developed and followed?

Sometimes the terms "implementation evaluation" and "monitoring evaluation" are confused. They are not the same. An implementation evaluation is an early check by the project staff, or the evaluator, to see if all essential elements are in place and operating. Monitoring is an external check. The monitor typically comes from the funding agency and is responsible for determining progress and compliance on a contract or grant for the project. Although the two differ, implementation evaluation, if effective, can facilitate project implementation and ensure that there are no unwelcome surprises during monitoring.

The purpose of a progress evaluation is to assess progress in meeting the goals.

Progress Evaluation. The purpose of a progress evaluation is to assess progress in meeting the goals of the program and the project. It involves collecting information to learn whether or not the benchmarks of participant progress were met and to point out unexpected developments. Progress evaluation collects information to determine what the impact of the activities and strategies is on participants, curriculum, or institutions at various stages of the intervention. By measuring progress, program staff can eliminate the risk of waiting until participants have experienced the entire program to assess likely outcomes. If the data collected as part of the progress evaluation fail to show expected changes, the information can be used to fine tune the project. Data collected as part of a progress evaluation can also contribute to, or form the basis for, a summative evaluation conducted at some future date. In a progress evaluation of the LSAMP program, the following questions can be addressed:

- Are the participants moving toward the anticipated goals of the project? Are they enhancing their academic skills? Are they gaining confidence in themselves as successful learners? Are they improving their understanding of the research process?

- Are the numbers of students reached increasing? How do changes in project participation relate to changes in the overall enrollments in mathematics, science, and technology areas at

their institutions? Are students being retained in their programs at an increasing rate?

- Does student progress seem sufficient in light of the long range goals of the program and project to increase the number of traditionally underrepresented students who receive degrees in science, mathematics, or technology?

Progress evaluation is useful throughout the life of the project, but is most vital during the early stages when activities are piloted and their individual effectiveness or articulation with other project components is unknown.
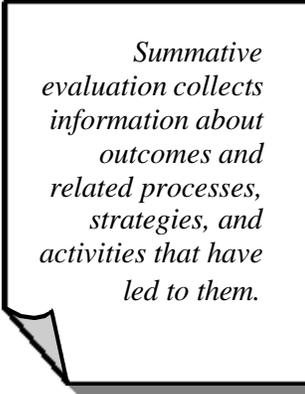
Summative Evaluation

The purpose of summative evaluation is to assess a mature project's success in reaching its stated goals. Summative evaluation (sometimes referred to as impact or outcome evaluation) frequently addresses many of the same questions as a progress evaluation, but it takes place after the project has been established and the timeframe posited for change has occurred. A summative evaluation of an LSAMP project might address these basic questions:
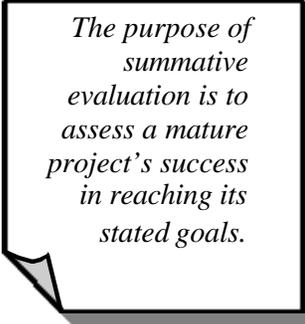
The purpose of summative evaluation is to assess a mature project's success in reaching its stated goals.

- To what extent does the project meet the stated goals for change or impact?

- Are greater numbers of students from diverse backgrounds receiving bachelor's of science degrees and showing increased interest in scientific careers?

- Are there any impacts on the schools participants attend? Are there any changes in courses? Are there any impacts of the LSAMP program on overall course offering and support services offered by their institution(s)?

- Which components are the most effective? Which components are in need of improvement?

- Were the results worth the program's cost?

- Can the program be sustained?

- Is the program replicable and transportable?

Summative evaluation collects information about outcomes and related processes, strategies, and activities that have led to them.
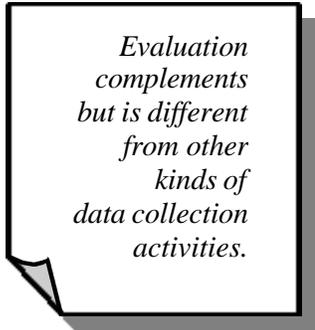
Summative evaluation collects information about outcomes and related processes, strategies, and activities that have led to them. The evaluation is an appraisal of worth, or merit. Usually this type of evaluation is needed for decisionmaking. The decision alternatives may include the following: disseminate the intervention to other sites or agencies; continue funding; increase funding; continue on probationary status; modify and try again; and discontinue.

In most situations, especially high-stakes situations or situations that are politically charged, it is important to have an external evaluator who is seen as objective and unbiased. Appendix A provides some tips for finding an evaluator. If this is not possible, it is better to have an internal evaluation than none at all. One compromise between the external and the internal model is to conduct an internal evaluation and then hire an outside agent to both review the design and assess the validity of the findings and conclusions.

When conducting a summative evaluation, it is important to consider unanticipated outcomes. These are findings that emerge during data collection or data analyses that were never anticipated when the study was first designed. For example, consider an NSF program providing professional development activities for teacher leaders. An evaluation intended to assess the extent to which participants share their new knowledge and skills with their school-based colleagues might uncover a relationship between professional development and attrition from the teaching force. These results could suggest new requirements for participants or cautions to bear in mind.

Evaluation Compared to Other Types of Data Gathering Activities

It is useful to understand how evaluation complements, but may differ from, other types of data collection activities that provide information on accountability for an NSF-funded project. Exhibit 4 shows various types of data collection activities, each of which provides somewhat different information and serves somewhat differing purposes. The continuum includes descriptive statistics, performance indicators, formative evaluation, summative evaluation, and research studies.
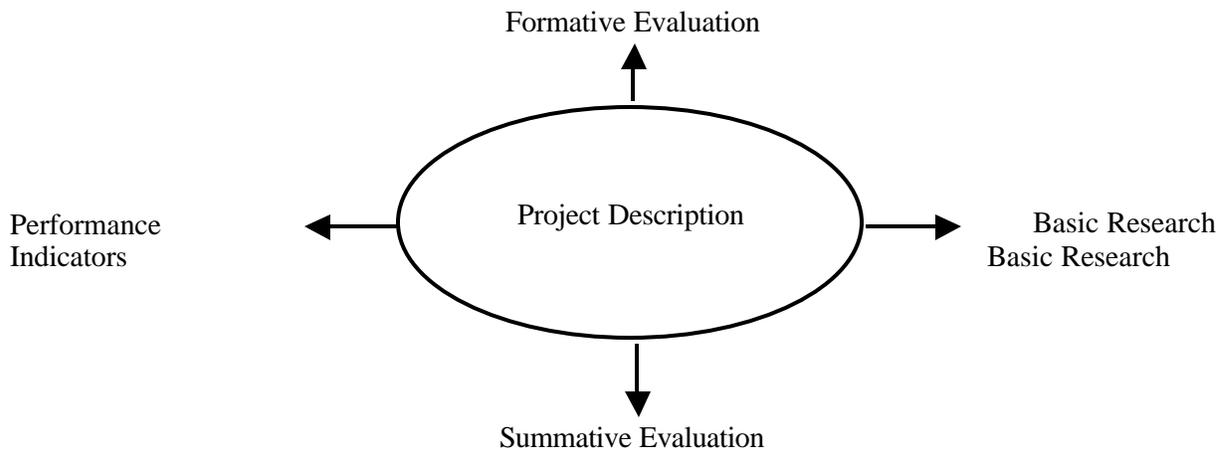
Evaluation complements but is different from other kinds of data collection activities.

At the center of the effort is the project description, which provides general information about a project. These data are commonly used to monitor project activities (e.g., funding levels, total number of participants), to describe specific project components (e.g., duration of program activity, number of participants enrolled in each activity), and to identify the types of individuals receiving services. Descriptive information may be collected annually or even more frequently to

provide a basic overview of a project and its accomplishments. Obtaining descriptive information usually is also part of each of the other data gathering activities depicted. NSF has developed the FASTLANE system as one vehicle for collecting such statistics.

FASTLANE allows for basic data to be collected across all programs in a consistent and systematic fashion. In addition, some programs have added program-specific modules aimed at collecting tailored data elements.

Exhibit 4.—Types of data gathering activities

Formative Evaluation

Performance Indicators ← Project Description → Basic Research Basic Research

Summative Evaluation

Formative and summative evaluations are intended to gather information to answer a limited number of questions. Evaluations include descriptive information, but go well beyond that. Generally, formative and summative evaluations include more indepth data collection activities, are intended to support decisionmaking, and are more costly.

Performance indicators fall somewhere between general program statistics and formative/summative evaluations. A performance indicator system is a collection of statistics that can be used to monitor the ongoing status of a program against a set of targets and metrics. Going beyond descriptive statistics, performance indicators begin to provide information that can be measured against a set of goals and objectives. Indicator systems are typically used to focus policymakers, educators, and the public on (1) key aspects of how an educational program is operating, (2) whether progress is being made, and (3) where there are problems (Blank, 1993). Because performance indicators focus on tangible results, they often go beyond traditional reviews of program expenditures and activity levels. In fact, the term "performance" underscores the underlying purpose of indicator systems, i.e., to examine a program's accomplishments and measure progress toward specific

goals. Performance indicators provide a snapshot of accomplishments in selected areas; however, in contrast to evaluations, the information is limited and is unlikely to provide an explanation of why a project may have succeeded or failed.

Research studies include descriptive information and provide targeted indepth exploration of issues, but differ along other dimensions. Instead of being intended for decisionmaking, research efforts typically are designed to explore conceptual models and alternative explanations for observed relationships.

Summary

The goal of evaluation is to determine the worth or merit of some procedure, project, process, or product. Well-designed evaluations also provide information that can help explain the findings that are observed. In these days of reform, educators are continually faced with the challenges of evaluating their innovations and determining whether progress is being made or stated goals have, in fact, been reached. Both common sense and accepted professional practice would suggest a systematic approach to these evaluation challenges. The role that evaluation may play will vary depending on the timing, the specific questions to be addressed, and the resources available. It is best to think of evaluation not as an event, but as a process. The goal should be to provide an ongoing source of information that can aid decisionmaking at various steps along the way.

References

Blank, R. (1993) Developing a System of Education Indicators: Selecting, Implementing, and Reporting Indicators. *Educational Evaluation and Policy Analysis*, 15 (1, Spring): 65-80.